8-12-2016

Some Novel Statistical Inferences

Chenxue Li

SOME NOVEL STATISTICAL INFERENCES

by

CHENXUE LI

Under the Direction of Gengsheng Qin and Liang Peng

ABSTRACT

In medical diagnostic studies, the area under the Receiver Operating Characteristic (ROC) curve (AUC) and Youden index are two summary measures widely used in the evaluation of the diagnostic accuracy of a medical test with continuous test results. The first half of this dissertation will highlight ROC analysis including extension of Youden index to the partial Youden index as well as novel confidence interval estimation for AUC and Youden index in the presence of covariates in induced linear regression models. Extensive simulation results show that the proposed methods perform well with small to moderate sized samples. In addition, some real examples will be presented to illustrate the methods.

The latter half focuses on the application of empirical likelihood method in economics and finance. Two models draw our attention. The first one is the predictive regression model with independent and identically distributed errors. Some uniform tests have been proposed in the literature without distinguishing whether the predicting variable is stationary or n-early integrated. Here, we extend the empirical likelihood methods in Zhu, Cai and Peng (2014 [1]) with independent errors to the case of an AR error process. The proposed new tests do not need to know whether the predicting variable is stationary or nearly integrated, and whether it has a finite variance or an infinite variance. Another model we considered is a GARCH(1,1) sequence or an AR(1) model with ARCH(1) errors. It is known that the observations have a heavy tail and the tail index is determined by an estimating equation. Therefore, one can estimate the tail index by solving the estimating equation with unknown parameters replaced by Quasi Maximum Likelihood Estimation (QMLE), and profile empirical likelihood method can be employed to effectively construct a confidence interval for the tail index. However, this requires that the errors of such a model have at least finite fourth moment to ensure asymptotic normality with $\sqrt{n}$ rate of convergence and Wilk's Theorem. We show that the finite fourth moment can be relaxed by employing some Least Absolute Deviations Estimate (LADE) instead of QMLE for the unknown parameters by noting that the estimating equation for determining the tail index is invariant to a scale transformation of the underlying model. Furthermore, the proposed tail index estimators have a normal limit with $\sqrt{n}$ rate of convergence under minimal moment condition, which may have an infinite fourth moment, and Wilk's theorem holds for the proposed profile empirical likelihood methods. Hence a confidence interval for the tail index can be obtained without estimating any additional quantities such as asymptotic variance.

INDEX WORDS: ROC Analysis, Partial Youden Index, GPQ, MOVER, AR Errors, Empirical Likelihood, Jackknife Empirical Likelihood, GARCH Sequence, Tail Index.

SOME NOVEL STATISTICAL INFERENCES

by

CHENXUE LI

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2016

SOME NOVEL STATISTICAL INFERENCES

by

CHENXUE LI

Committee Chair:	Gengsheng Qin
	Liang Peng
Committee:	Ruiyan Luo
	Xin Qi

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

August 2016

DEDICATION

To my beloved parents, my advisors,

and all my dear friends!

ACKNOWLEDGEMENTS

This dissertation could not be finished without the help and love from many people I have met in the Department of Mathematics and Statistics at Georgia State University. I'd like to acknowledge and dedicate my gratitude to the following people who generously shared their care and time with me.

First of all, I would like to express my special appreciation and thanks to my respected advisors Professor Gengsheng Qin and Professor Liang Peng. Thank you for being such great and tremendous mentors. During my five years' study abroad, you have taken me in as a member of your families and directed me to the optimal path of my life in every aspect. Your advice on both research as well as on my career have been priceless. Dr. Qin has spent tremendous effort to make me better. All the lessons I have learned from our discussions and talks will benefit me for the rest of my career. I cannot imagine how lucky I was to have Dr. Peng as my co-chair and advisor as well. He is the person who inspired me to explore the world of business and dive into a brand new area which I have come to love. Thank you both for giving me the opportunity to be one of your students, which makes me so proud to be here.

Here, I would also like to thank Professor Zehui Li and Dr. Jinyuan Chen from Lanzhou University, who were two very important people to me during my school years. I've received their endless care and encouragement not just during my undergraduate study in China, but also when I was doing my graduate program at Georgia State University. They cheered me up when I felt frustrated and showed their support whenever I needed it. I collaborated with Dr. Chen for almost 2 years and he has been extremely patient and provided impressive insights.

Next, my appreciation goes to my committee members, Professor Ruiyan Luo and Professor Xin Qi for serving as my committee members. You witnessed my growth from immature to mature. I still remember the questions you asked 3 years ago, and now I have more complete answers. Thank you for equipping me with the foundation of statistics by taking your classes.

Additionally, I cannot forget any of the professors I have met here who have taught and helped me with my teaching and study at Georgia State University, Dr. Guantao Chen, Dr. Alexandra Smirnova, Dr. Florian Enescu, Dr. Draga Vidakovic, Dr. Valerie Miller, Dr. Leslie Meadows, Dr. Sutandra Sarkar, Dr. Yichuan Zhao, Dr. Frank Hall, Dr. Zhongshan Li, Dr. Changyong Zhong, Dr. Samara Marko, Dr. Xu Zhang, Dr. Jun Han, Dr. Yuanhui Xiao as well as others: Sandra Ahuama-Jonas, Earnestine Collier, Beth Conner. Thank you for all your help in the past 5 years.

A special thanks to my family and my friends. My words cannot express how grateful I am to my parents for all of the sacrifices that they have made on my behalf. Their unconditional love is my motivation towards achieving my goals. Even with small progress, they could not be happier. My life would not be so colorful without my friends and colleagues: Haiqi Wang, Binhuan Wang, Haochuan Zhou, Jie Zhang, Yanan Yin, Zhujun Li, Jun Xia, Bing Liu, Xiaolin Li, Siyu Tian, Shuenn Siang Ng, Wei Gao, Chuanyun Zang, Jing Wang, Guangming Jing, Junwei Wang, Howard Smith, Robin Baidya. Last but not the least, I'd like to thank my boyfriend Benjamin Kim and his sweet friends, who are more experienced in job hunting and have helped me analyze current job market. My success today is credited to all my friends' understanding and love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

- GSU - Georgia State University
- ROC - Receiver Operating Characteristic
- AUC - Area Under the Curve
- pAUC - Partial Area Under the Curve
- YI - Youden Index
- pYI - Partial Youden Index
- GPQ - Generalized Pivotal Quantities
- HAC - Hybrid Agresti-Coull
- HWS - Hybrid Wilson-Score
- SHWS - Symmetric Hybrid Wilson-Score
- MOVER - Method Of Variance Estimates Recovery
- EL - Empirical Likelihood
- JEL - Jackknife Empirical Likelihood
- GARCH - Generalized Autoregressive Conditional Heteroskedasticity
- AR - Autoregressive
- ARMA - Autoregressive Moving Average
- ARCH - Autoregressive Conditional Heteroskedasticity
- QMLE - Quasi Maximum Likelihood Estimation

- LADE - Least Absolute Deviations Estimate
- PEL - Profile Empirical Likelihood

Chapter 1

INTRODUCTION

1.1 Statistical Evaluation of Medical Tests

In modern medicine, diagnostic/screening tests are essential procedures to prevent, detect and treat diseases. An accurate test can provide reliable information about the condition of subjects under diagnosis and influence the plan of the test users for managing the subjects.

Biopsy tests with high accuracy are often understood as the reliable diagnostic methods. However, the costs of biopsy tests including extreme pain, tissue removal, neuron damage and operational costs cannot be ignored. A compromise is to make diagnosis based on alternative tests (e.g., biomarkers, body symptoms) with acceptable diagnostic accuracy.

Diagnostic errors always exist. For a test with binary outcomes, a subject will be classified into either a healthy group or a diseased group based on it's test result. False negative (*FN*) error that refers to classifying a diseased individual as non-diseased, and false positive (*FP*) error that refers to classifying a non-diseased individual as diseased, are two types of errors resulting from the inaccuracy of a test (Pierce, 1884 [2]). False Positive Rate (*FPR*), False Negative Rate (*FNR*), True Negative Rate (*TNR*, *1-FPR*, also called specificity), and True Positive Rate (*TPR*,*1-FNR*, also called sensitivity), are commonly used parameters for measuring the accuracy of a test.

When outcomes of a test are binary, specificity and sensitivity of the test can be calculated. When outcomes of a test are continuous, selection of a threshold/cut-off ("c") point is necessary to define the positivity of test results. Let random variable X denote the test result from the non-diseased group, and random variable Y denote the test result from the diseased group. Without loss of generality, assume that $Y > X$. Then, for a given cut-off point c, specificity $= P(X \leq c)$, and sensitivity $= P(Y > c)$. In order to evaluate the overall performance of a test, we'd better take all possible c into account. A plot of sensitivity vs.

1-specificity over all the possible cut-off points is called the "Receiver Operating Characteristic (ROC) curve". A perfect diagnostic test has an ROC curve starting from the origin, going straight to (0, 1), then turning right at ninety degrees and ending at (1, 1).

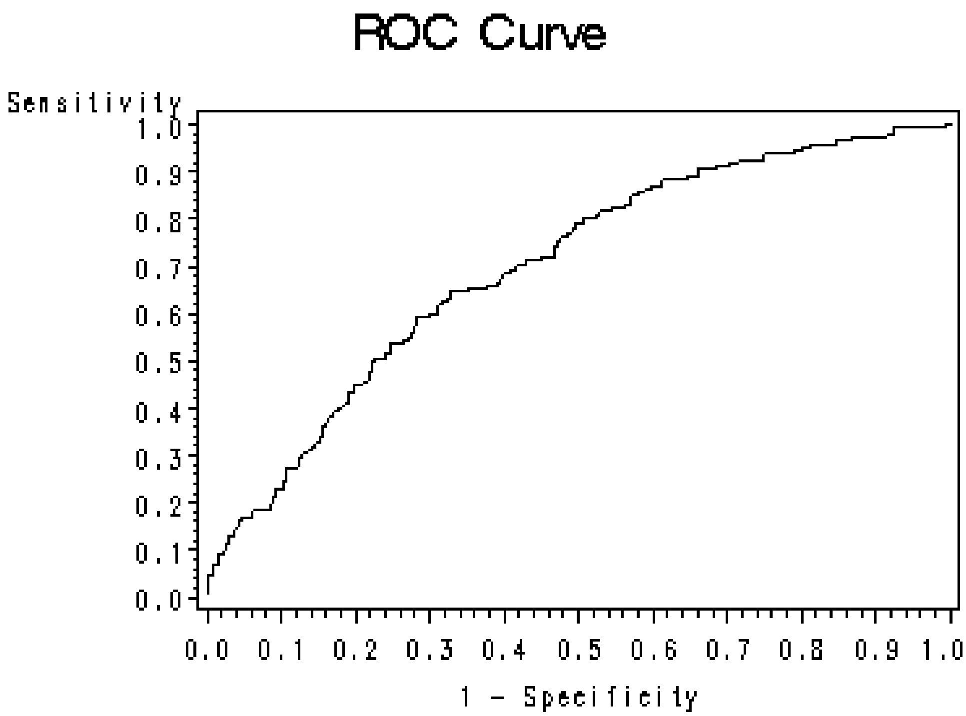


Figure (1.1) Example of ROC curve [3]

Different diagnostic/screening tests may have different ROC curves. The test with higher sensitivity and specificity is a favorable test. However, sometimes the curves may cross. Consequently, it causes a problem as we compare performances of tests by merely looking at the ROC curves. One way to solve this problem is to evaluate the "Area Under the ROC curve" (AUC) of the test. AUC is a "one number summary" of an ROC curve. Bamber (1975 [4]) proved that $AUC = P(Y > X)$. In the ideal situation, AUC has a value of 1. AUC has been used in a wide range of scientific fields such as signal detection theory, medical imaging, weather forecasting, and diagnostic medicine. Some references on inferences of AUC were provided by Refs. ([5], [6], [7]).

Although AUC has many advantages in summarizing the accuracy of a diagnostic test, it has its limitations. In some circumstances, ROC curve might be used to represent test performance on a truncated range of clinically relevant values of FPR, or if one wished to exclude those parts of the ROC space where study data are sparse [8]. The ROC curve extends beyond the clinically relevant area of potential clinical interpretation. Hence, the concept of the partial AUC ($pAUC$) was proposed in literature. The partial AUC analysis has been recognized and many methods for inference on the partial AUC have been developed. With proper bi-normal model checking, McClish ([9], [10]) provided a method for comparing portion of ROC curves. Based on McClish's work, Jiang et al. ([11]) proposed a partial area index for highly sensitive diagnostic tests (Zhang et al. [6]). Dodd and Pepe (2003 [12]) gave very good interpretations for $pAUC$.

1.1.1 Inference for Youden Index

The AUC and the partial AUC are widely used summary measures for an ROC curve. But they can not be used to select a cut-off level with desired sensitivity/specificity. A wise choice of a cut-off point is an important implementation for a test. Several selection methods of cut-off point including "CB" (cost-benefit method), "MinValueSp" (a minimum value set for Specificity), "MinValueSe" (a minimum value set for Sensitivity), "RangeSp" (a range of values set for Specificity), and "RangeSe" (a range of values set for Sensitivity) have been proposed and can be found in R program package 'Optimal Cutpoints' (Miller and Siegmund, 1982 [13], Altman, et al., 1994 [14]).

Here we focus on another "one number summary" of the ROC curve, Youden index (J), which is defined as follows:

$$J = \max_c \{sensitivity(c) + specificity(c) - 1\} \tag{1.1}$$

$$= sensitivity(c_0) + specificity(c_0) - 1 \tag{1.2}$$

where c_0 is the optimal cut-point of the test.

Youden index was first introduced by Youden [15] in 1950. Indubitably, both high sensitivity and specificity are desired for a medical test. Schisterman and Perkins [16] pointed out that the optimal threshold for the positive test result of a disease should be the threshold leading to the maximum of the sum of TPR and TNR. At the same time, this optimal cut-off point also guarantees minimization of the sum of FPR and FNR, and Youden's index illustrates this simply and clearly. Youden index (J) represents the maximum differentiating ability of a biomarker when equal weight is given to sensitivity and specificity, with J ranging from 0 to 1 where 0 indicates the test has no discriminating ability and 1 indicates the test is perfect (Fluss et al., 2005 [17]). It not only supplies a method to find an optimal cut-off point, but also provides a numerical summary of the classification likelihood of the test. From a graphical perspective, Youden's index is the maximum vertical distance between the ROC curve and the diagonal chance line, which is in accord with the differentiating capacity of the diagnosis. This index has several remarkable features, such as it is independent of the relative/absolute sizes of the diseased and non-diseased groups, and all tests that share the same index make the same total number of misclassifications per hundred patients (Youden 1950 [15]).

As mentioned above, the ROC curve is constructed by plotting "1-specificity" against "sensitivity" at all possible cut-off points. Let X denote the test result from a non-diseased population with distribution $F(x)$ and $\{X_i : i = 1, 2, \ldots, n\}$ is a random sample from $F(x)$. Let Y denote the test result from diseased population with distribution $G(y)$ and $\{Y_j : j = 1, 2, \ldots, m\}$ is a random sample from $G(y)$. For given cut-off point c, we have

$$sensitivity(c) = P(Y \geq c) = 1 - G(c) \tag{1.3}$$

$$specificity(c) = P(X < c) = F(c). \tag{1.4}$$

Youden index can be written as follows:

$$J = \max_c \{1 - G(c) + F(c) - 1\} \tag{1.5}$$

$$= \max_c \{F(c) - G(c)\} \tag{1.6}$$

$$= F(c_0) - G(c_0) \tag{1.7}$$

where c_0 is the optimal cut-point of the test results.

Hsieh and Turnbull [18] studied non-parametric estimation methods for the Youden index based on the empirical and kernel estimates for the underlying distributions. They provided asymptotic properties of the estimators. However, the asymptotic variances for the empirical estimate of Youden's index is still unknown, thus confidence intervals for the Youden index cannot be constructed directly. Some studies (e.g., Faraggi [19]) considered constructing non-parametric confidence intervals for the Youden index and the corresponding cutoff point. Zhou and Qin [20] focused on construction of non-parametric confidence intervals for the Youden index and provided two new non-parametric intervals for the Youden index based on Agresti and Coull's [21] adjusted estimate for a binomial proportion.

In practice, high sensitivity (e.g., $0.90 < sensitivity(c) < 1$) or high specificity (e.g., $0.8 < specificity(c) < 1$, [22]) is of special interest for a medical test. However, no method for finding the Youden index along with corresponding cut-off point on a partial interval of possible cut-off points has been proposed. Inspired by the motivation for the partial AUC, we will propose a new summary index, called "partial Youden index" on a partial interval of possible cut-off points for a continuous-scale test. The traditional Youden index is a special case of the proposed partial Youden index. More details on the partial Youden index and its inference will be discussed in Chaper 2.

1.1.2 AUC and Youden Index in the Presence of Covariates

Nowadays, modern medical services enable us to collect more information of our patients. The extra information about the individual, other than the test result, for instance, the age, the gender, the race etc. are called "covariates". Ignoring the information on covariates may cause low accuracy of the diagnostic/screening test. The following is an example to show that incorporating age as a covariate significantly enhances the accuracy of the test.

It is a population-based cross-sectional pilot survey of diabetes mellitus in Cairo, Egypt, and consists of postprandial blood glucose measurements of 286 subjects obtained from a fingerstick (a clip of the data set 1.1.2). According to the gold standard criteria of the World

Health Organization for diagnosing diabetes, 88 subjects were classified as diseased and 198 subjects as healthy. The age of the subject was considered as a relevant covariate in this example because glucose levels are expected to be higher for older people who do not suffer from diabetes (see Ref[23], for details).

finger	gold	age
82	0	37
82	0	20
87	1	51
80	0	54

The following table lists several measures of accuracy calculated from the same test with and without consideration of the covariate "age"("age" is linked to test results by a linear model).

covariate	AUC	J	TPR	TNR
without	0.9057	0.6718	0.8894	0.7824
with	0.9543	0.7768	0.9654	0.8115

Apparently, with the consideration of "age", the accuracy of the test has been improved, which is very beneficial to our diagnosis.

Four general types of covariate information can be incorporated into the models for diagnostic accuracy study: 1) subject characteristics, 2) clinical indicators, 3) confounding variables, and 4) test operating parameters([24]). Dodd [24] pointed out that these 4 types of covariate information are not mutually exclusive.

In order to evaluate the influence of covariates on AUC and Youden index, some researchers have used induced-regression methods. They modeled the test/biomarker values through regression models in each population separately. Pepe [25] and Tosteson et al. [26] specified models for test results as a function of disease status and covariates. Smith and Thompson [23] proposed a parametric survival model for modeling the distribution of the screening test outcome as a function of true disease status and other confounding covariates.

Zhou et al. [27] extended the models proposed in Pepe [25] by allowing for heteroscedasticity. Zheng and Heagerty [28] proposed a semi-parametric estimator for the conditional ROC curve, in which the distribution of the error terms is unknown and allowed to depend on the covariates, but, as in the previous articles, the effect of the covariates on the conditional means and variances is modeled parametrically. Recently, Rodríguez and Martínez [29] presented a Bayesian semi-parametric model, in which the error terms are assumed to be normally distributed, but non-parametric specifications of the conditional means and variances are allowed.

The models mentioned above are generally complex. Faraggi [30] used simple linear regression to model biomarker values from the diseased and non-diseased populations. These linear regression models permit the examination of a direct connection between covariates and biomarker values within each population. Using maximum likelihood estimation and the normal approximation method, Faraggi [30] obtained an adjusted confidence interval for the AUC. Faraggi [30] also provided adjusted confidence intervals for the Youden index and the corresponding critical threshold value by using a bootstrap method. However, these methods may not perform well with small sample sizes. In addition, the bootstrap-based methods are computationally time-consuming.

1.2 Uniform Test Predictive Regression Models

Predictive regression models have been widely used in economics and finance. A simple predictive regression model goes as follows:

$$Y_t = \alpha + \beta X_{t-1} + U_t, \quad X_t = \theta + \phi X_{t-1} + e_t, \quad B(L)e_t = V_t, \tag{1.8}$$

where $L^i e_t = e_{t-i}$, $B(L) = 1 + (\sum_{i=1}^{q} b_i L^i)$, $B(1) \neq 1$, all the roots of $B(L)$ are fixed and less than one in absolute value, and $(U_1, V_1), ..., (U_n, V_n)$ are independent and identically distributed random vector with mean zero and finite variances. Testing $H_0 : \beta = 0$ or constructing a confidence interval for β answers the important question whether the variable

X_{t-1} can be used to predict Y_t.

When $\{X_t\}$ is a stationary sequence, the least squares estimator for β based on the first equation in (1.8) can be used to formulate a simple test and to construct a confidence interval. However, this simple estimator ignores the dependence between U_t and V_t and so tends to be biased in finite sample.This motivates the study of proposing some bias-corrected estimators and tests in the literature; see Stambaugh (1999 [31]), Amihud and Hurvich (2004 [32]), Amihud, Hurvich and Wang (2009 [33]), Chen and Deo (2009 [34]).

When the predicting variable X_{t-1} is a macroeconomic variable such as the log dividend-price ratio or the log earnings-price ratio, the assumption of stationarity for $\{X_t\}$ is quite questionable. On the other hand, it is known that the asymptotic limit of the simple least squares estimator is quite different when X_t has a finite variance or an infinite variance, and when the sequence $\{X_t\}$ is stationary or nearly integrated. Therefore, having a unified test or interval estimation is of importance in practice, which avoids the extremely challenging tasks in detecting whether the sequence $\{X_t\}$ is stationary or nearly integrated, and whether X_t has a finite variance or an infinite variable. Some existing uniform tests proposed in the literature include the Bonferroni t-test of Cavanagh, Elliott and Stock (1995 [35]), the Bonferroni Q-test of Campbell and Yogo (2006 [36]), and the empirical likelihood test of Zhu, Cai and Peng (2014 [1]). All these methods assume that $U_t's$ are independent and identically distributed random variables, which may be quite restrictive in practice, which draws our attention.

1.3 Tail Index of GARCH(1,1) Model and AR(1) Model with ARCH(1) Errors

A large number of empirical studies show that many financial data series, such as exchange rate returns and stock indices, often exhibit skewness and heavy tails (see Taylor (2005 [37])). The heaviness of tails determines some unusual asymptotic behavior of sample covariance functions, sample correlation functions and extremes of the underlying sequence; see Davis and Resnick (1985, 1986 [38] [39]) for ARMA processes, Mikosch and Stărică (2000 [40]) and Basrak, Davis and Mikosch (2002 [41]) for GARCH sequences, Davis and Mikosch

(1998 [42]) and De Haan, Resnick, Rootzén and de Vries (1989 [43]) for ARCH models, Borkovec (2000, 2001 [44] [45]) for an AR(1) process with ARCH(1) errors, and Davis and Resnick (1996 [46]) and Resnick and Van den Berg (2000 [47]) for bilinear time series. When the sequence follows from a time series model, heavy tailed errors play an important role in deriving the asymptotic limit of parameters estimation; see Hall and Yao (2003 [48]) for the study of quasi maximum likelihood estimation (QMLE) for a GARCH process, Lange (2011 [49]) and Zhang and Ling (2015 [50]) for the study of least squares estimation for AR-GARCH models. Some robust inference procedures for heavy-tailed GARCH models can be found in Hill (2015 [51]) and Hill and Prokhorov (2016 [52]). The tail index also plays an important role in testing structural changes in stock prices (see Quinton, Fan and Phillips (2001 [53])) and calculating financial risk measures such as Value-at-Risk and expected shortfall (see Wagner and Marsh (2005 [54])). Therefore inference for the tail index is useful in understanding and modeling time series data.

For a GARCH(p,q) sequence, i.e.,

$$Y_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = w + \sum_{i=1}^{p} a_i \sigma_{t-i}^2 + \sum_{j=1}^{q} b_j Y_{t-j}^2,$$

where $w > 0, a_i \geq 0, b_i \geq 0$ are unknown parameters, and $\varepsilon_t' s$ are independent and identically distributed random variables with zero mean and variance one, Basrak, Davis and Mikosch (2002 [41]) showed that, under some conditions, there is $\alpha > 0$ such that $\lim_{x\to\infty} x^\alpha P(|Y_t| > x) \in (0, \infty)$ by using results in Kesten (1973 [55]) for random difference equations. For estimating the tail index α, one could simply employ the Hill's estimator (see Hill (1975)) defined as

$$\tilde{\alpha}(k) = \{\frac{1}{k} \sum_{i=1}^{k} \log \frac{Y_{n,n-i+1}}{Y_{n,n-k}}\}^{-1}, \tag{1.9}$$

where $Y_{n,1} < \cdots < Y_{n,n}$ denote the order statistics of $Y_1, \cdots, Y_n$, and $k = k(n) \to \infty$ and $k/n \to 0$ as $n \to \infty$. Although the Hill's estimator has been studied extensively for independent data, existing research on dependent data such as m-dependence or β-mixing can be found in Hsing (1991 [56]), Resnick and Stărică (1998 [57]) and Drees (2000 [58]). However the choice of the sample fraction k for dependent data is much more complicated than that for independent data. Indeed, as far as we are aware, there is no data-driven method for choosing k for dependent data although several methods are available for independent data. For the case of an ARMA sequence, one can apply the Hill's estimator to either the sequence itself or the estimated errors since heavy tailed errors imply that the observations have a heavy tail with the same tail index (see Resnick and Stărică (1997 [59]) and Ling and Peng (2004 [60])).

When $p = q = 1$, i.e., a GARCH(1,1) sequence, the tail index α is determined by an estimating equation (see Section 5.2 for details). In this case, the tail index can be estimated by using all observations rather than a small fraction of upper order statistics as Hill's estimator, and so the resulted estimator has a faster rate of convergence than the Hill's estimator and does not need to choose the sample fraction k. Asymptotic limit was first derived in Mikosch and Stărică (2000 [40]), and later its asymptotic variance was corrected by Berkes, Horváth and Kokoszka (2003 [61]). Since the asymptotic variance is very complicated, interval estimation relies on bootstrap method, which is computationally extensive due to the fact that one has to resample from estimated errors and refit the GARCH(1,1) model. Moreover, it is known that the performance of bootstrap method for non-pivotal statistics is not good in general. Therefore Chan, Peng and Zhang (2012 [62]) proposed a profile empirical likelihood method to construct a confidence interval for the index α by using score equations derived from QMLE, which requires finite fourth moment for ϵ_t. For an AR(1) model with ARCH(1) errors, which is sometimes called a double AR process in the literature, the tail index is determined by an estimating equation too. Therefore Chan, Li, Peng and Zhang (2013 [63]) derived the asymptotic limit of an estimator for the index based on an estimation equation with QMLE, and proposed a profile empirical likelihood method to construct a confidence interval without estimating the asymptotic variance explicitly. These results require the errors to have a finite fourth moment as well.

1.4 Aims of the Dissertation

As mention in section 1.1, no method for finding the Youden index along with corresponding cut-off point on a partial interval of possible cut-off points has been proposed. Our first aim is to extend the traditional Youden index to a more generalized index called "partial Youden index". In addition, we will propose various parametric and non-parametric methods for inference on the partial Youden index including Generalized Confidence Interval Method, Method of Variance Estimation Recovery (Hybrid Wilson-Score, Hybrid Agresti-Coull, and Symmetric Hybrid Wilson-Score). Simulation studies will be conducted to evaluate the performance of the proposed methods.

Considering the significance of incorporating covariate into test results, in Chapter 3, we chose linear models to model the covariate and the test results. Our second aim is to provide an exact "generalized confidence interval" for the AUC and Youden index. The new method is going to be compared with some existing methods.

The third aim of this dissertation research is to investigate the possibility of extending the unified approach in Zhu, Cai and Peng (2014 [1]) to the case in which $\{U_t\}$ follows an AR(p) process. When $\{U_t\}$ is an α-mixing sequence, estimation and test are proposed by Cai and Wang (2014 [64]), which do not lead to a unified procedure.

Motivated by the analysis of the exchange rates between Hong Kong dollar and US dollar in Zhu and Ling (2015 [65]), our fourth aim is to propose a robust method to estimate the tail index, which allows the errors to have an infinite fourth moment. More specifically, by noting that the estimating equation for determining the tail index is invariant to a scale transformation of the studied models, we propose to first estimate the unknown parameters by a least absolute deviations estimate (LADE) and then to estimate the tail index by the estimating equation. This leads to a tail index estimator with the $\sqrt{n}$ rate of convergence and asymptotic normality without requiring a finite fourth moment of errors. Since the asymptotic variance of the proposed tail index estimator is too complicated, we further propose to employ the profile empirical likelihood method to construct a confidence interval, which does not require to estimate the asymptotic variance explicitly. Unlike existing methods in Berkes, Horváth and Kokoszka (2003 [61]), Chan, Peng and Zhang (2012 [62]) and Chan, Li, Peng and Zhang (2013 [66]), the proposed methods not only relax the moment conditions of errors (see Section 5.2), but also perform well because LADE is more robust than QMLE (see the empirical study in Section 5.3).

1.5 Outline of the Dissertation

The remainder of this dissertation is organized as follows. Chapters 2 and 3 are concentrated on "Inference for Partial Youden Index" and "Inference for AUC and Youden Index with Covariate Adjustment", respectively. Chapters 4 and 5 can be categorized as the application of Empirical Likelihood method and its derivatives for two different time series problems, which include "Uniform Test for Predictive Regression with AR errors" and "Inference for Tail Index of GARCH(1,1) and AR(1) Model with ARCH(1) Errors Under Minimal Moment Condition". All the proofs are provided in the Appendices.

Chapter 2

INFERENCE FOR PARTIAL YOUDEN INDEX

2.1 Review and Outline

In section 1.1, we have introduced the definition of Youden index. From formula (1.2), we can see that Youden index is a function of sensitivity and specificity depending on the underlying distributions of the diseased and non-diseased populations. Many methods have been proposed for inference on Youden index. Most of them need assumptions about their underlying distributions (e.g., binormal distributions). Fluss et al. [17] proposed parametric estimate for Youden index. Schisterman and Perkins [16] provided parametric and non-parametric confidence intervals for the index.

In this chapter, we will first define a "partial Youden index" (pYI) for a medical test. Both parametric and non-parametric methods will be proposed to construct confidence intervals for the partial Youden index using Generalized Pivotal Quantities (GPQs, see Weerahandi [67]) and "Method of Variance Estimates Recovery" ("MOVER") (Zou and Donner, 2008 [68]). Extensive simulation studies will be conducted to evaluate the finite sample performances of the new intervals. At last, our proposed method will be applied to a real problem for comparing the diagnostic accuracy of two biomarkers ("CA-125" vs. "CA-19-9") for the detection of pancreatic cancer.

2.2 Motivation

As mentioned in Chapter 1, in diagnostic studies, high sensitivity (e.g., $0.90 <$ sensitivity(c) < 1) or high specificity (e.g., $0.8 <$ specificity(c) < 1, [22]) is of special interest for a medical test. Dating back to 1989, the partial area under the ROC curve was first proposed by McClish (1989 [9]), Thompson and Zucchini (1989 [69]). Dodd [12] pointed out that the partial AUC was an alternative measure to the full AUC. When using the partial AUC, one considers only those regions of the ROC space where data have been observed, or which correspond to clinically relevant values of test sensitivity or specificity [12]. Inspired by the motivation for the partial AUC, we propose a new summary index, called "partial Youden index" on a partial interval (c_2, c_1) of possible cut-off points for a continuous-scale test as follows:

$$J_{p_1,p_2} = \max_{c_2 \le c \le c_1} \{sensitivity(c) + specificity(c) - 1\} \tag{2.1}$$

$$= sensitivity(c_{po}) + specificity(c_{po}) - 1 \tag{2.2}$$

$$= F(c_{po}) - G(c_{po}) \tag{2.3}$$

where (p_1, p_2) $(0 \le p_1 < p_2 \le 1)$ is an interval of FPRs of interest such that $c_1 = F^{-1}(1-p_1)$ and $c_2 = F^{-1}(1-p_2)$, c_{po} is the optimal cut-off point corresponding to the partial Youden Index.

Remarks:

1. If $p_1 = 0$, $p_2 = 1$, then $c_1 = \infty$, $c_2 = -\infty$, the partial Youden index is reduced to the Youden index on the full interval of cut-off points.

2. J_{p_1,p_2} defined above is the partial Youden index with restriction on specificity. If (p_1, p_2) $(0 \le p_1 < p_2 \le 1)$ is an interval of TPRs of interest such that $c_1 = G^{-1}(1-p_1)$ and $c_2 = G^{-1}(1-p_2)$, then J_{p_1,p_2} is the partial Youden index with restriction on sensitivity. For simplicity, we will only consider the partial Youden index with restriction on specificity in this chapter.

2.3 Methodologies

Generalized Confidence Interval In the following, we will briefly review the basic concept of the generalized confidence interval proposed by Weerahandi (1993 [67]).

Suppose that Y is a random variable whose distribution depends on (θ, δ), where θ is a parameter of interest and δ is a nuisance parameter. Let y be the observed value of Y. A generalized pivotal quantity (GPQ) $R(Y; y, \theta, \delta)$, a function of Y, y, θ, and δ, for interval estimation, defined in Weerahandi (1993 [67]), satisfies the following conditions:

(1) $R(Y; y, \theta, \delta)$ has a distribution free of all unknown parameters.

(2) The value of $R(Y; y, \theta, \delta)$ at $Y = y$ is θ, the parameter of interest.

To derive a GPQ-based confidence for the partial Youden index, we assume that X and Y are independent, and the underlying distributions $F(x)$ and $G(y)$ of the non-diseased and diseased populations are $N(\mu_x, \sigma_x^2)$ and $N(\mu_y, \sigma_y^2)$, respectively. Without loss of generality, we assume that $\mu_x < \mu_y$.

The following point estimates for the Youden index along with its optimal cut-off point were given in Schisterman and Perkins (2007 [16]):

$$c_0 = \frac{\mu_x(b^2 - 1) - a + b\sqrt{a^2 + (b^2 - 1)\sigma_x^2 \ln b^2}}{b^2 - 1} \tag{2.4}$$

and

$$J = \Phi\left(\frac{\mu_y - c_0}{\sigma_y}\right) + \Phi\left(\frac{c_0 - \mu_x}{\sigma_x}\right) \tag{2.5}$$

where $a = \mu_y - \mu_x, b = \frac{\sigma_y}{\sigma_x}$, and $\Phi(\cdot)$ denotes the standard normal cumulative distribution function.

When $\sigma_x^2 = \sigma_y^2$, c_0 is undefined and it can be replaced by

$$c_0 = \frac{\mu_x + \mu_y}{2} \tag{2.6}$$

which is the limit of (2.4) as $b \to 1$.

For the "non-diseased" and the "diseased" samples $\{X_i : i = 1, \cdots, n\}$ and $\{Y_j : j = 1, \cdots, m\}$, let $\bar{X}$, $\bar{Y}$ be the sample means and S_x^2, S_y^2 be the sample variances. Let $\bar{x}$, $\bar{y}$, s_x^2 and s_y^2 be the observed values of $\bar{X}$, $\bar{Y}$, S_x^2 and S_y^2, respectively. The generalized pivotal quantities of μ_x and μ_y are

$$R_{\mu_x} = \bar{x} - \left(\frac{\bar{X} - \mu_x}{\sigma_x/\sqrt{n}}\right)\frac{\sigma_x}{S_x}\frac{s_x}{\sqrt{n}} = \bar{x} - \frac{Z_x}{\sqrt{V_x/(n-1)}}\frac{s_x}{\sqrt{n}} = \bar{x} - t_x\frac{s_x}{\sqrt{n}} \tag{2.7}$$

$$R_{\mu_y} = \bar{y} - \left(\frac{\bar{Y} - \mu_y}{\sigma_y/\sqrt{m}}\right)\frac{\sigma_y}{S_y}\frac{s_y}{\sqrt{m}} = \bar{y} - \frac{Z_y}{\sqrt{V_y/(m-1)}}\frac{s_y}{\sqrt{m}} = \bar{y} - t_y\frac{s_y}{\sqrt{m}}, \tag{2.8}$$

respectively, where $Z_x = \frac{\sqrt{n}(\bar{X}-\mu_x)}{\sigma_x} \sim N(0,1), Z_y = \frac{\sqrt{m}(\bar{Y}-\mu_y)}{\sigma_y} \sim N(0,1), V_x = \frac{(n-1)S_x^2}{\sigma_x^2} \sim \chi^2_{n-1}, V_y = \frac{(m-1)S_y^2}{\sigma_y^2} \sim \chi^2_{m-1}$ and $t_x = \frac{Z_x}{\sqrt{V_x/(n-1)}}, t_y = \frac{Z_y}{\sqrt{V_y/(m-1)}}$ follow Student's t-distribution with degrees of freedom $n-1$ and $m-1$, respectively.

The generalized pivotal quantities for σ_x^2 and σ_y^2 are given by

$$R_{\sigma_x^2} = \frac{\sigma_x^2}{(n-1)S_x^2}(n-1)s_x^2 = \frac{(n-1)s_x^2}{V_x}, \tag{2.9}$$

$$R_{\sigma_y^2} = \frac{\sigma_y^2}{(m-1)S_y^2}(m-1)s_y^2 = \frac{(m-1)s_y^2}{V_y}, \tag{2.10}$$

respectively.

The generalized pivotal quantities for σ_x and σ_y are $R_{\sigma_x} = \sqrt{R_{\sigma_x^2}}$ and $R_{\sigma_y} = \sqrt{R_{\sigma_y^2}}$, respectively.

The GPQs for a and b are

$$R_a = R_{\mu_y} - R_{\mu_x}, \quad R_b = \frac{R_{\sigma_y}}{R_{\sigma_x}},$$

respectively.

Therefore, the GPQs for c_0 and J are

$$R_{c_0} = \frac{R_{\mu_x}(R_b^2 - 1) - R_a + R_b\sqrt{R_a^2 + (R_b^2 - 1)R_{\sigma_x^2}\ln R_b^2}}{R_b^2 - 1}, \tag{2.11}$$

when the variances are equal,

$$R_{c_0} = \frac{R_{\mu_x} + R_{\mu_y}}{2}, \tag{2.12}$$

$$R_J = \Phi\left(\frac{R_{\mu_y} - R_{c_0}}{R_{\sigma_y}}\right) + \Phi\left(\frac{R_{c_0} - R_{\mu_x}}{R_{\sigma_x}}\right) - 1, \tag{2.13}$$

respectively.

Based on the definition of the partial Youden index, we can consider the following three situations to find the optimal cut-off point c_{po} for the partial Youden Index:

1. If the regular optimal cut-off point c_0 for the Youden index is located between c_2 and c_1, then $c_{po} = c_0$.

2. If the regular optimal cut-off point c_0 for the Youden index is located to the left side of c_2, then $c_{po} = c_2$.

3. If the regular optimal cut-off point c_0 for the Youden index is located to the right side of c_1, then $c_{po} = c_1$.

Hence, c_{po} = median of (c_0, c_1, c_2). To construct the generalized confidence interval for the partial Youden Index, we need to derive the GPQ for the optimal cut-off point c_{po}.

Note that

$$c_1 = F^{-1}(1 - p_1), \quad c_2 = F^{-1}(1 - p_2). \tag{2.14}$$

Since $F(x)$ is $N(\mu_x, \sigma_x^2)$,

$$c_1 = \sigma_x \Phi^{-1}(1 - p_1) + \mu_x, \quad c_2 = \sigma_x \Phi^{-1}(1 - p_2) + \mu_x. \tag{2.15}$$

Consequently, the GPQs for c_1 and c_2 are $R_{c_1} = R_{\sigma_x}\Phi^{-1}(1-p_1)+R_{\mu_x}$ and $R_{c_2} = R_{\sigma_x}\Phi^{-1}(1-p_2) + R_{\mu_x}$, respectively. Therefore, the GPQ for c_{po} is $R_{c_{po}} = median(R_{c_0}, R_{c_1}, R_{c_2})$.

The following algorithm is proposed to construct the generalized confidence interval (GCI) for the partial Youden Index.

Algorithm:

For given “non-diseased” and “diseased” samples $x_1, \ldots, x_n$ and $y_1, \ldots, y_m$,

1. Compute the sample means $\bar{x}$ and $\bar{y}$ and sample variances s_x^2 and s_y^2.

2. For $k = 1, \ldots, K$

   - Generate t_{n-1} and t_{m-1};
   - Generate V_x and V_y from χ^2_{n-1} and χ^2_{m-1}, respectively;
   - Compute $R_{\mu_x}, R_{\mu_y}, R_{\sigma_x}$, and R_{σ_y};
   - Compute R_{c_0}, R_{c_1};
   - Compute $R_{c_{po},k} = median(R_{c_0}, R_{c_1}, R_{c_2})$;
   - Compute $R_{J_{p_1,p_2},k}$ by replacing R_{c_0} by $R_{c_{po},k}$ in R_J.

   (end k loop)

3. Compute the $100\alpha/2$-th percentile $R_{J_{p_1,p_2},\alpha/2}$ and the $100(1-\alpha/2)$-th percentile $R_{J_{p_1,p_2},(1-\alpha)/2}$ of $\{R_{J_{p_1,p_2},1}, R_{J_{p_1,p_2},2}, \ldots, R_{J_{p_1,p_2},K}\}$. Then, $\left(R_{J_{p_1,p_2},\alpha/2}, R_{J_{p_1,p_2},(1-\alpha)/2}\right)$ is a $100(1-\alpha)\%$level confidence interval for J_{p_1,p_2}.

4. Compute the $100\alpha/2$th percentile $R_{c_{po},\alpha/2}$ and the $100(1-\alpha/2)$th percentile $R_{c_{po},(1-\alpha)/2}$ of $\{R_{c_{po},1}, R_{c_{po},2}, \ldots, R_{c_{po},K}\}$. Then, $\left(R_{c_{po},\alpha/2}, R_{c_{po},(1-\alpha)/2}\right)$ is a $100(1-\alpha)\%$ level confidence interval for c_{po}.

Non-Parametric Hybrid Confidence Intervals The generalized confidence interval has its limitation because it's a parametric interval. Alternatively, non-parametric method can also be considered. Here, we employ the "Method of Variance Estimates Recovery" ("MOVER") (Zou and Donner, 2008 [68]) and square-and-add approach (Newcombe, 1998 [70]) to construct non-parametric hybrid confidence intervals for the partial Youden index.

From

$$
\begin{aligned}
J_{p_1,p_2} &= \max_{c_2 \le c \le c_1} \{sensitivity(c) + specificity(c) - 1\} \\
&= sensitivity(c_{po}) + specificity(c_{po}) - 1 \\
&= sensitivity(c_{po}) - (1 - specificity(c_{po})) \\
&= P(Y \ge c_{po}) - P(X \ge c_{po}) \\
&\equiv \theta_2 - \theta_1,
\end{aligned}
$$

we can see that the partial Youden index is the difference between two unknown proportions θ_2 and θ_1, where $\theta_1 \equiv P(X \ge c_{po})$, $\theta_2 \equiv P(Y \ge c_{po})$. However, the two proportions can estimated by $\hat{\theta}_1 = \sum_{i=1}^{n} I(X_i \ge \hat{c}_{po})/n$ and $\hat{\theta}_2 = \sum_{j=1}^{m} I(Y_j \ge \hat{c}_{po})/m$, where $\hat{c}_{po}$ is a consistent estimate for c_{po} (e.g., the empirical estimate for c_{po}). Hence, $\hat{J}_{p_1,p_2} = \hat{\theta}_2 - \hat{\theta}_1$ is a consistent estimate for the partial Youden index.

Under the assumption that the test results from non-diseased group and diseased group are independent, the variance of $\hat{J}_{p_1,p_2}$ can be consistently estimated by

$$
\widehat{Var}(\hat{J}_{p_1,p_2}) = \widehat{Var}(\hat{\theta}_2 - \hat{\theta}_1) = \widehat{Var}(\hat{\theta}_2) + \widehat{Var}(\hat{\theta}_1)
$$

where $\widehat{Var}(\hat{\theta}_1) = \hat{\theta}_1(1 - \hat{\theta}_1)/n$ and $\widehat{Var}(\hat{\theta}_2) = \hat{\theta}_2(1 - \hat{\theta}_2)/m$ are consistent estimates for the variance of $\hat{\theta}_1$ and $\hat{\theta}_2$, respectively.

The $(1 - \alpha)$-th Wald-Type confidence interval for the partial Youden index is:

$$
\left(\hat{J}_{p_1,p_2} - z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\theta}_1) + \widehat{Var}(\hat{\theta}_2)}, \hat{J}_{p_1,p_2} + z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\theta}_1) + \widehat{Var}(\hat{\theta}_2)} \right) \tag{2.16}
$$

Our simulation studies showed that this Wald-Type confidence interval has poor small sample performance. In order to improve the performance of the Wald-type CI, we use the MOVER method (See also Zou et al. 2009 [71]) to construct new hybrid confidence intervals for the partial Youden index.

Let l_i and u_i $(i = 1, 2)$ be the lower and upper limits of a $100(1-\alpha)\%$ two-sided CI for θ_i. From the Central Limit Theorem, it follows that

$$l_i = \hat{\theta}_i - z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\theta}_i)},$$

$$u_i = \hat{\theta}_i + z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\theta}_i)},$$

which implies that the variance of $\hat{\theta}_i$ can be estimated by $\widehat{Var}_l(\hat{\theta}_i) = (\hat{\theta}_i - l_i)^2/z^2_{\alpha/2}$ and $\widehat{Var}_u(\hat{\theta}_i) = (u_i - \hat{\theta}_i)^2/z^2_{\alpha/2}$. After plugging these variance estimates back to equation (2.16), we get the following hybrid confidence intervals for the partial Youden index:

$$\left(\hat{J}_{p_1,p_2} - \sqrt{(\hat{\theta}_2 - l_2)^2 + (u_1 - \hat{\theta}_1)^2},\ \hat{J}_{p_1,p_2} + \sqrt{(u_2 - \hat{\theta}_2)^2 + (\hat{\theta}_1 - l_1)^2}\right).$$

Here, we propose the following methods to get two-sided CI (l_i, u_i) for θ_i.

(i) The Agresti-Coull method.

$$l_1 = \tilde{\theta}_1 - z_{\alpha/2}\sqrt{\frac{\tilde{\theta}_1(1-\tilde{\theta}_1)}{n + z^2_{\alpha/2}}}, \qquad u_1 = \tilde{\theta}_1 + z_{\alpha/2}\sqrt{\frac{\tilde{\theta}_1(1-\tilde{\theta}_1)}{n + z^2_{\alpha/2}}}$$

where $\tilde{\theta}_1 = (\sum_{i=1}^{n} I(X_i \geq \hat{c}_{po}) + 0.5z^2_{\alpha/2})/(n + z^2_{\alpha/2})$.

$$l_2 = \tilde{\theta}_2 - z_{\alpha/2}\sqrt{\frac{\tilde{\theta}_2(1-\tilde{\theta}_2)}{m + z^2_{\alpha/2}}}, \qquad u_2 = \tilde{\theta}_2 + z_{\alpha/2}\sqrt{\frac{\tilde{\theta}_2(1-\tilde{\theta}_2)}{m + z^2_{\alpha/2}}}$$

where $\tilde{\theta}_2 = (\sum_{j=1}^{m} I(Y_j \geq \hat{c}_{po}) + 0.5z^2_{\alpha/2})/(m + z^2_{\alpha/2})$.

The hybrid confidence interval based on this method is called Hybrid Agresti-Coull (HAC) interval for the partial Youden index. It can be seen that the *HAC* interval is a symmetric interval.

(ii) The Wilson score method.

$$l_1 = \frac{\hat{\theta}_1 + \frac{z^2_{\alpha/2}}{2n} - z_{\alpha/2}\sqrt{\frac{\hat{\theta}_1(1-\hat{\theta}_1)}{n} + \frac{z^2_{\alpha/2}}{4n^2}}}{1 + z^2_{\alpha/2}/n}, \qquad u_1 = \frac{\hat{\theta}_1 + \frac{z^2_{\alpha/2}}{2n} + z_{\alpha/2}\sqrt{\frac{\hat{\theta}_1(1-\hat{\theta}_1)}{n} + \frac{z^2_{\alpha/2}}{4n^2}}}{1 + z^2_{\alpha/2}/n}.$$

$$l_2 = \frac{\hat{\theta}_2 + \frac{z^2_{\alpha/2}}{2m} - z_{\alpha/2}\sqrt{\frac{\hat{\theta}_2(1-\hat{\theta}_2)}{m} + \frac{z^2_{\alpha/2}}{4m^2}}}{1 + z^2_{\alpha/2}/m}, \qquad u_2 = \frac{\hat{\theta}_2 + \frac{z^2_{\alpha/2}}{2m} + z_{\alpha/2}\sqrt{\frac{\hat{\theta}_2(1-\hat{\theta}_2)}{m} + \frac{z^2_{\alpha/2}}{4m^2}}}{1 + z^2_{\alpha/2}/m}$$

The hybrid confidence interval based on this method is called Hybrid Wilson Score (HWS) interval for the partial Youden index. Shan [72] proposed two improved confidence intervals for Youden index using the square-and-add limits based on the Wilson score method. Shan's method is equivalent to the above Hybrid Wilson Score method.

(iii) The Symmetric Hybrid Wilson Score

We also can construct a symmetric interval based on the Hybrid Wilson Score method. If $l_{HWS} = \hat{J}_{p_1,p_2} - \Delta_l$ and $u_{HWS} = \hat{J}_{p_1,p_2} + \Delta_u$ are the lower and upper limits of $100(1-\alpha)\%$ two-sided CI for J_{p_1,p_2} based on the Hybrid Wilson Score method. A Symmetric Hybrid Wilson Score (SHWS) confidence interval for the partial Youden index is defined as

$$(l_{SHWS},\ u_{SHWS}),$$

where

$$l_{SHWS} \equiv \hat{J}_{p_1,p_2} - \sqrt{(\Delta_l^2 + \Delta_u^2)/2}, \qquad u_{SHWS} \equiv \hat{J}_{p_1,p_2} + \sqrt{(\Delta_l^2 + \Delta_u^2)/2}.$$

2.4 Simulation Studies

To evaluate the finite sample performance of the proposed method, two simulation studies are conducted to compare the coverage probabilities ("cp") and average lengths ("al") of the GPQ-based interval, the Hybrid Agresti-Coull (HAC) interval, the Hybrid Wilson-Score (HWS)/Shan's [72](NP) intervals, and the Symmetric Hybrid Wilson-Score

(SHWS) interval when the underlying distributions are bi-normal distributions and Gamma distributions, respectively. When the underlying distributions are Gamma distributions, GPQ-based method can not be applied. In the simulation study, we apply the Box-Cox transformation to the simulated data from the Gamma distributions, and then calculate the coverage probabilities and average lengths of the GPQ-based intervals based on the transformed data.

In the first simulation study, we generate the "non-diseased" sample $\{x_i : i = 1, \cdots, n\}$ from a normal distribution with mean $\mu_x = 0$ and variance $\sigma_x^2 = 1$ and the "diseased" sample $\{y_j : j = 1, \cdots, m\}$ from a normal distribution with mean μ_y and variances $\sigma_y^2 = 3$, respectively. In the second simulation study, the "non-diseased" sample is generated from $Gamma(1.5, 1)$, and the "diseased" sample is generated from $Gamma(2, \theta_y)$, where the values for μ_y and θ_y are calculated such that the true Y-ouden index $J = 0.5, 0.8$, respectively. In the two studies, we choose sample sizes $(n, m) = (20, 20), (20, 40), (40, 20), (40, 40), (80, 80)$, respectively, $K = 1000$, $(p_1, p_2) = (0, 0.01), (0, 0.1), (0.01, 0.2), (0.05, 0.1), (0.05, 0.2), (0.1, 0.3), (0, 1)$, where $(p_1, p_2) = (0, 0.01)$ represents the case with extremely high specificity, and $(p_1, p_2) = (0, 1)$ represents the case for the traditional Youden index which is a special case of the partial Youden index. 1000 iterations were made to compute the coverage probabilities and average lengths of the 95% confidence intervals.

Table (3.1)- (3.2) display the simulation results when the underlying distributions of the test results are bi-normal distributions. Table (3.3)- (3.4) display the results when the underlying distributions of the test results are Gamma distributions.

From Table (3.1)- (3.2), we observe that GPQ method shows its prominent performance consistently with both the sample sizes and the selected values of (p_1, p_2). All the coverage probabilities of the GPQ-based intervals are close to 95% nominal level. For the non-parametric HAC, $HWS(NP)$, and $SHWS$ intervals, their coverage probabilities vary for different combinations of (p_1, p_2). When $p_2 - p_1$ is small, these non-parametric interval-s show under-coverage problem. When $p_2 - p_1$ increases, their coverage probabilities also increase. As the sample sizes m and n get bigger, these non-parametric intervals perform better. Performances of HAC and $HWS(NP)$ methods also are related to the true values of the Youden index. When the true Youden index is small, HAC and HWS intervals perform well. When the true Youden index is big, HWS confidence intervals often overestimate the partial Youden index. The coverage probabilities of HWS are far greater than the nominal level. Another observation is that $SHWS$ intervals sometimes fail when $J = 0.8$ (See Table (3.2)). HAC intervals have coverage probabilities closer to 95% than the $HWS(NP)$ intervals.

From Table (3.3)-(3.4), we can see that the HAC and $HWS(NP)$ methods stand out among all the methods. Similar to the bi-normal case, the non-parametric intervals perform better when the sample sizes m and n get bigger, especially for HAC intervals. When $J = 0.5$, GPQ-based interval is not recommended, but when $J = 0.8$, GPQ-based intervals have acceptable performances (see Table (3.4)). Similar to those results for the bi-normal distributions, when $J = 0.8$, $SHWS$ intervals fails sometimes, and HAC intervals have coverage probabilities closer to 95% than the $HWS(NP)$ intervals.

In summary, we recommend the GPQ-based interval when the underlying distributions follow bi-normal distributions, and the non-parametric HAC interval when the underlying distributions are unknown.

2.5 A Real Example

In this section, we apply our methods to a dataset on pancreatic cancer. The dataset are the outcomes of two biomarkers "CA-125" and "CA-19-9", which include tests results from 51 "control" patients and 90 "case" patients.

Wieand et al. [22] plotted the ROC curves of "CA-125" and "CA-19-9", and demonstrated that there were some differences between the two curves when the specificity falls in (0.8, 1) [73]. This motivates us to focus on this interval to evaluate the diagnostic ability of the biomarkers in terms of the partial Youden index.

Specificity within (0.8, 1) corresponds to $p_1 = 0$ and $p_2 = 0.2$. Since the original data

Table (2.1) The coverage probabilities and average lengths of the 95% confidence interval for the pYI in Normal Case $J = 0.5$

m	n	(p_1, p_2)	pYI	GPQ		HAC		$HWS(NP)$		$SHWS$	
				cp	al	cp	al	cp	al	cp	al
20	20	(0,0.01)	0.349	0.957	0.470	0.915	0.469	0.880	0.466	0.791	0.473
		(0,0.1)	0.496	0.959	0.445	0.951	0.475	0.945	0.465	0.897	0.478
		(0.01,0.2)	0.5	0.962	0.420	0.974	0.479	0.957	0.467	0.903	0.483
		(0.05,0.1)	0.496	0.963	0.443	0.960	0.475	0.954	0.466	0.920	0.479
		(0.05,0.2)	0.5	0.954	0.420	0.973	0.479	0.945	0.467	0.901	0.483
		(0.1,0.3)	0.5	0.954	0.412	0.975	0.488	0.964	0.476	0.921	0.493
		(0,1)	0.5	0.956	0.412	0.987	0.481	0.961	0.471	0.909	0.487
20	40	(0,0.01)	0.349	0.957	0.413	0.894	0.415	0.894	0.414	0.822	0.416
		(0,0.1)	0.496	0.951	0.394	0.958	0.425	0.940	0.421	0.873	0.427
		(0.01,0.2)	0.5	0.965	0.383	0.979	0.426	0.931	0.421	0.888	0.429
		(0.05,0.1)	0.496	0.952	0.395	0.957	0.426	0.924	0.422	0.856	0.429
		(0.05,0.2)	0.5	0.956	0.382	0.973	0.428	0.949	0.423	0.906	0.431
		(0.1,0.3)	0.5	0.953	0.378	0.978	0.432	0.950	0.425	0.903	0.435
		(0,1)	0.5	0.961	0.378	0.975	0.424	0.950	0.418	0.895	0.428
40	20	(0,0.01)	0.349	0.961	0.419	0.902	0.390	0.842	0.385	0.768	0.394
		(0,0.1)	0.496	0.962	0.373	0.957	0.399	0.949	0.389	0.909	0.403
		(0.01,0.2)	0.5	0.946	0.345	0.972	0.407	0.948	0.395	0.894	0.409
		(0.05,0.1)	0.496	0.962	0.372	0.945	0.399	0.938	0.389	0.907	0.403
		(0.05,0.2)	0.5	0.959	0.345	0.975	0.408	0.960	0.397	0.909	0.411
		(0.1,0.3)	0.5	0.952	0.337	0.989	0.419	0.978	0.407	0.928	0.421
		(0,1)	0.5	0.961	0.340	0.989	0.412	0.968	0.401	0.906	0.415
40	40	(0,0.01)	0.349	0.957	0.343	0.849	0.322	0.825	0.322	0.755	0.323
		(0,0.1)	0.496	0.961	0.312	0.949	0.340	0.942	0.337	0.907	0.341
		(0.01,0.2)	0.5	0.950	0.295	0.956	0.345	0.932	0.340	0.889	0.346
		(0.05,0.1)	0.496	0.946	0.310	0.952	0.342	0.944	0.339	0.919	0.343
		(0.05,0.2)	0.5	0.948	0.295	0.953	0.347	0.932	0.343	0.893	0.348
		(0.1,0.3)	0.5	0.949	0.292	0.965	0.354	0.955	0.349	0.915	0.355
		(0,1)	0.5	0.956	0.293	0.963	0.346	0.936	0.341	0.884	0.347
80	80	(0,0.01)	0.349	0.949	0.246	0.791	0.222	0.775	0.222	0.775	0.222
		(0,0.1)	0.496	0.943	0.217	0.938	0.241	0.921	0.240	0.889	0.242
		(0.01,0.2)	0.5	0.952	0.208	0.939	0.248	0.925	0.246	0.889	0.248
		(0.05,0.1)	0.496	0.960	0.219	0.945	0.244	0.932	0.243	0.915	0.245
		(0.05,0.2)	0.5	0.949	0.208	0.944	0.250	0.936	0.248	0.900	0.250
		(0.1,0.3)	0.5	0.953	0.207	0.964	0.254	0.957	0.252	0.920	0.254
		(0,1)	0.5	0.953	0.208	0.953	0.249	0.940	0.247	0.901	0.249

Table (2.2) The coverage probabilities and average lengths of the 95% confidence interval for the pYI in Normal Case $J = 0.8$

m	n	(p_1, p_2)	pYI	GPQ		HAC		$HWS(NP)$		$SHWS$	
				cp	al	cp	al	cp	al	cp	al
20	20	(0,0.01)	0.727	0.956	0.461	0.941	0.422	0.969	0.393	0.933	0.428
		(0,0.1)	0.8	0.951	0.300	0.957	0.401	0.987	0.364	0.997	0.409
		(0.01,0.2)	0.8	0.950	0.294	0.978	0.401	0.993	0.364	0.997	0.409
		(0.05,0.1)	0.8	0.946	0.293	0.957	0.402	0.980	0.365	0.992	0.410
		(0.05,0.2)	0.8	0.955	0.284	0.971	0.402	0.992	0.365	0.998	0.410
		(0.1,0.3)	0.792	0.952	0.249	0.964	0.413	0.988	0.380	1.000	0.420
		(0,1)	0.8	0.960	0.289	0.975	0.399	0.997	0.361	1.000	0.407
20	40	(0,0.01)	0.727	0.953	0.383	0.962	0.370	0.941	0.351	0.869	0.373
		(0,0.1)	0.8	0.953	0.270	0.973	0.343	0.984	0.314	0.915	0.347
		(0.01,0.2)	0.8	0.954	0.266	0.983	0.338	0.987	0.308	0.926	0.343
		(0.05,0.1)	0.8	0.945	0.264	0.964	0.346	0.979	0.318	0.924	0.351
		(0.05,0.2)	0.8	0.958	0.256	0.980	0.340	0.989	0.310	0.926	0.345
		(0.1,0.3)	0.792	0.945	0.232	0.963	0.350	0.985	0.325	0.998	0.355
		(0,1)	0.8	0.956	0.267	0.985	0.339	0.991	0.310	0.940	0.344
40	20	(0,0.01)	0.727	0.954	0.412	0.945	0.351	0.956	0.327	0.899	0.357
		(0,0.1)	0.8	0.960	0.247	0.972	0.339	0.981	0.309	0.985	0.345
		(0.01,0.2)	0.8	0.944	0.240	0.976	0.341	0.987	0.312	0.979	0.346
		(0.05,0.1)	0.8	0.945	0.243	0.961	0.340	0.966	0.311	0.968	0.346
		(0.05,0.2)	0.8	0.961	0.233	0.981	0.340	0.992	0.311	0.978	0.346
		(0.1,0.3)	0.792	0.954	0.206	0.983	0.358	0.993	0.332	1.000	0.362
		(0,1)	0.8	0.952	0.240	0.985	0.341	0.992	0.312	0.989	0.347
40	40	(0,0.01)	0.727	0.962	0.320	0.914	0.286	0.895	0.274	0.839	0.287
		(0,0.1)	0.8	0.939	0.207	0.973	0.268	0.969	0.250	0.915	0.270
		(0.01,0.2)	0.8	0.968	0.206	0.988	0.269	0.974	0.251	0.934	0.271
		(0.05,0.1)	0.8	0.953	0.203	0.953	0.273	0.970	0.255	0.952	0.275
		(0.05,0.2)	0.8	0.950	0.202	0.978	0.273	0.985	0.256	0.954	0.275
		(0.1,0.3)	0.792	0.956	0.176	0.984	0.283	0.995	0.267	1.000	0.284
		(0,1)	0.8	0.946	0.205	0.979	0.268	0.970	0.250	0.920	0.270
80	80	(0,0.01)	0.727	0.946	0.225	0.872	0.197	0.829	0.193	0.766	0.197
		(0,0.1)	0.8	0.964	0.146	0.975	0.185	0.972	0.178	0.917	0.185
		(0.01,0.2)	0.8	0.944	0.145	0.978	0.184	0.964	0.177	0.894	0.185
		(0.05,0.1)	0.8	0.940	0.145	0.969	0.188	0.968	0.181	0.939	0.188
		(0.05,0.2)	0.8	0.953	0.145	0.982	0.188	0.985	0.181	0.945	0.188
		(0.1,0.3)	0.792	0.951	0.126	0.985	0.195	0.990	0.189	0.982	0.196
		(0,1)	0.8	0.945	0.145	0.976	0.184	0.955	0.177	0.904	0.185

Table (2.3) The coverage probabilities and average lengths of the 95% confidence interval for the pYI in Gamma Case $J = 0.5$

m	n	(p_1, p_2)	pYI	*GPQ*		*HAC*		*HWS(NP)*		*SHWS*	
				cp	al	cp	al	cp	al	cp	al
20	20	(0,0.01)	0.242	0.771	0.431	0.805	0.460	0.767	0.459	0.685	0.465
		(0,0.1)	0.465	0.872	0.513	0.930	0.476	0.939	0.468	0.826	0.480
		(0.01,0.2)	0.5	0.912	0.478	0.965	0.481	0.939	0.469	0.859	0.486
		(0.05,0.1)	0.465	0.848	0.501	0.938	0.477	0.941	0.468	0.822	0.481
		(0.05,0.2)	0.5	0.896	0.480	0.950	0.481	0.928	0.469	0.877	0.486
		(0.1,0.3)	0.5	0.900	0.443	0.973	0.485	0.948	0.471	0.882	0.489
		(0,1)	0.5	0.913	0.415	0.984	0.481	0.950	0.468	0.885	0.486
20	40	(0,0.01)	0.242	0.725	0.356	0.847	0.400	0.774	0.399	0.774	0.402
		(0,0.1)	0.465	0.843	0.440	0.936	0.428	0.925	0.424	0.856	0.430
		(0.01,0.2)	0.5	0.890	0.411	0.957	0.427	0.912	0.420	0.859	0.430
		(0.05,0.1)	0.465	0.847	0.440	0.924	0.430	0.928	0.427	0.860	0.432
		(0.05,0.2)	0.5	0.879	0.417	0.963	0.430	0.924	0.423	0.880	0.433
		(0.1,0.3)	0.5	0.910	0.389	0.979	0.429	0.941	0.421	0.900	0.432
		(0,1)	0.5	0.898	0.365	0.973	0.421	0.935	0.411	0.866	0.423
40	20	(0,0.01)	0.242	0.878	0.426	0.772	0.384	0.737	0.381	0.652	0.389
		(0,0.1)	0.465	0.890	0.455	0.885	0.400	0.878	0.390	0.826	0.403
		(0.01,0.2)	0.5	0.917	0.408	0.959	0.414	0.945	0.402	0.885	0.416
		(0.05,0.1)	0.465	0.911	0.460	0.881	0.401	0.887	0.392	0.847	0.404
		(0.05,0.2)	0.5	0.913	0.405	0.953	0.412	0.927	0.401	0.875	0.415
		(0.1,0.3)	0.5	0.905	0.378	0.975	0.422	0.952	0.410	0.892	0.424
		(0,1)	0.5	0.921	0.358	0.989	0.417	0.960	0.406	0.887	0.419
40	40	(0,0.01)	0.242	0.797	0.332	0.746	0.311	0.702	0.312	0.702	0.312
		(0,0.1)	0.465	0.869	0.372	0.913	0.343	0.891	0.339	0.860	0.344
		(0.01,0.2)	0.5	0.911	0.333	0.955	0.349	0.926	0.344	0.879	0.350
		(0.05,0.1)	0.465	0.869	0.369	0.904	0.344	0.894	0.341	0.872	0.345
		(0.05,0.2)	0.5	0.899	0.328	0.933	0.351	0.914	0.345	0.867	0.351
		(0.1,0.3)	0.5	0.885	0.306	0.958	0.354	0.923	0.348	0.881	0.355
		(0,1)	0.5	0.923	0.299	0.958	0.350	0.932	0.345	0.872	0.351
80	80	(0,0.01)	0.242	0.824	0.253	0.703	0.210	0.679	0.210	0.665	0.210
		(0,0.1)	0.465	0.854	0.260	0.880	0.245	0.868	0.245	0.821	0.246
		(0.01,0.2)	0.5	0.904	0.228	0.923	0.253	0.923	0.251	0.889	0.253
		(0.05,0.1)	0.465	0.871	0.263	0.889	0.246	0.890	0.246	0.846	0.247
		(0.05,0.2)	0.5	0.920	0.225	0.936	0.253	0.928	0.251	0.887	0.253
		(0.1,0.3)	0.5	0.906	0.215	0.942	0.255	0.937	0.253	0.892	0.255
		(0,1)	0.5	0.909	0.214	0.957	0.254	0.952	0.251	0.903	0.254

Table (2.4) The coverage probabilities and average lengths of the 95% confidence interval for the pYI in Gamma Case $J = 0.8$

m	n	(p_1, p_2)	pYI	GPQ		HAC		$HWS(NP)$		$SHWS$	
				cp	al	cp	al	cp	al	cp	al
20	20	(0,0.01)	0.724	0.811	0.487	0.939	0.423	0.968	0.394	0.946	0.429
		(0,0.1)	0.8	0.912	0.341	0.956	0.400	0.984	0.362	0.998	0.408
		(0.01,0.2)	0.8	0.923	0.306	0.971	0.400	0.992	0.361	0.998	0.407
		(0.05,0.1)	0.8	0.935	0.341	0.963	0.401	0.988	0.363	0.995	0.409
		(0.05,0.2)	0.8	0.940	0.289	0.982	0.399	0.994	0.361	0.998	0.407
		(0.1,0.3)	0.793	0.944	0.254	0.960	0.414	0.992	0.380	1.000	0.420
		(0,1)	0.8	0.932	0.286	0.981	0.398	0.995	0.361	1.000	0.406
20	40	(0,0.01)	0.724	0.732	0.408	0.948	0.374	0.931	0.356	0.880	0.377
		(0,0.1)	0.8	0.915	0.304	0.967	0.343	0.982	0.314	0.923	0.347
		(0.01,0.2)	0.8	0.907	0.271	0.985	0.339	0.993	0.310	0.922	0.344
		(0.05,0.1)	0.8	0.931	0.294	0.975	0.344	0.990	0.315	0.929	0.348
		(0.05,0.2)	0.8	0.936	0.261	0.979	0.341	0.993	0.312	0.933	0.346
		(0.1,0.3)	0.793	0.941	0.228	0.974	0.347	0.988	0.321	0.999	0.352
		(0,1)	0.8	0.924	0.259	0.987	0.339	0.991	0.309	0.935	0.344
40	20	(0,0.01)	0.724	0.838	0.443	0.930	0.353	0.951	0.329	0.900	0.358
		(0,0.1)	0.8	0.928	0.279	0.967	0.339	0.979	0.309	0.974	0.344
		(0.01,0.2)	0.8	0.922	0.240	0.981	0.339	0.990	0.309	0.972	0.345
		(0.05,0.1)	0.8	0.939	0.264	0.975	0.340	0.983	0.310	0.984	0.345
		(0.05,0.2)	0.8	0.926	0.235	0.983	0.340	0.994	0.310	0.981	0.345
		(0.1,0.3)	0.793	0.936	0.204	0.971	0.358	0.990	0.333	1.000	0.362
		(0,1)	0.8	0.938	0.238	0.991	0.340	0.995	0.311	0.974	0.346
40	40	(0,0.01)	0.724	0.799	0.356	0.906	0.287	0.882	0.276	0.825	0.288
		(0,0.1)	0.8	0.921	0.220	0.967	0.269	0.967	0.251	0.933	0.271
		(0.01,0.2)	0.8	0.941	0.203	0.987	0.268	0.978	0.250	0.937	0.270
		(0.05,0.1)	0.8	0.934	0.217	0.958	0.273	0.971	0.255	0.952	0.275
		(0.05,0.2)	0.8	0.935	0.197	0.979	0.271	0.982	0.253	0.942	0.273
		(0.1,0.3)	0.793	0.945	0.178	0.984	0.282	0.991	0.266	1.000	0.284
		(0,1)	0.8	0.936	0.202	0.979	0.269	0.968	0.251	0.926	0.271
80	80	(0,0.01)	0.724	0.808	0.248	0.870	0.197	0.828	0.194	0.788	0.198
		(0,0.1)	0.8	0.908	0.148	0.970	0.184	0.961	0.177	0.908	0.185
		(0.01,0.2)	0.8	0.922	0.144	0.976	0.184	0.967	0.177	0.906	0.185
		(0.05,0.1)	0.8	0.936	0.145	0.974	0.187	0.974	0.180	0.937	0.188
		(0.05,0.2)	0.8	0.920	0.141	0.983	0.187	0.978	0.179	0.928	0.187
		(0.1,0.3)	0.793	0.932	0.126	0.983	0.195	0.988	0.188	0.996	0.195
		(0,1)	0.8	0.931	0.144	0.988	0.185	0.971	0.178	0.906	0.185

are not normally distributed, we use Box-Cox transformation with the power parameter $\phi = -0.425$ to the "CA-125" test results, and $\phi = -0.015$ to the "CA-19-9" tests results. Then the transformed data would follow normal distribution.

In order to make sure the stability of GPQ-based method, we choose $K = 10^6$ to construct 95% level GPQ-based interval for the partial Youden index $J_{0,0.2}$ based on the transformed data in this example. We also calculate 95% level HAC, HWS and $SHWS$ intervals for $J_{0,0.2}$ (See Table (2.5)).

Table (2.5) shows that "CA-19-9" has higher partial Youden index than "CA-125", which indicates that biomarker "CA-19-9" has higher diagnostic accuracy to detect pancreatic cancer than biomarker "CA-125" when specificity of the two biomarkers falls in (0.8, 1). Therefore, we recommend "CA-19-9" for detection of pancreatic cancer. This conclusion coincides with the results in Huang et al. [73]. Also, based on our proposed method, we can get the confidence interval for the optimal cut off point.

Table (2.5) The 95% confidence interval for the $J_{0,0.2}$ of CA-125 v.s. CA-19-9

	GPQ	HAC	HWS	$SHWS$
$CA-125$	(0.1167, 0.4012)	(0.1232, 0.4208)	(0.1296, 0.4259)	(0.1439, 0.4417)
$CA-19-9$	(0.5776, 0.7854)	(0.4944, 0.7434)	(0.5089, 0.7525)	(0.5328, 0.7823)

2.6 Summary and Discussion

In this chapter, we propose a new summary index, called "partial Youden index", for a ROC curve. We also develop parametric and non-parametric confidence intervals for the partial Youden index. The proposed methods are derived from GPQ and MOVER based methods. The partial Youden index maintains merits of the traditional Youden index. It can be a useful tool for finding an optimal cut-off point. In medial diagnostic studies, a test having minimum sensitivity or specificity is often required clinically. The proposed partial Youden index can assure a lower bound for sensitivity or specificity by adjusting the values of

p_1 and p_2. We also conduct extensive simulation studies to evaluate the proposed methods. Our simulation results show that the generalized confidence interval for the partial Youden index method performs very well when the underlying distributions are binormal. The non-parametric HAC interval has acceptable performance when the underlying distributions are unknown.

The partial Youden index is a new summary index for a ROC curve. The traditional Youden index is a special case of the proposed index. It is well known that Youden index has been applied to many fields in medical and biological studies. We expect the partial Youden index will have wilder applications in medical and biological sciences.

Chapter 3

INFERENCE FOR AUC AND YOUDEN INDEX WITH COVARIATE ADJUSTMENT

3.1 Review and Outline

As stated in Chapter 1.1.2, in this chapter, we consider similar linear regression models to those used in Faraggi [30] and provide method for constructing exact confidence intervals for AUC, and the Youden index along with its corresponding optimal cut-point. Our approach is based on the concept of generalized pivotal quantity (GPQ) introduced by Tsui [74] and Weerahandi [67] . When compared with the normal approximation-based intervals, the proposed generalized confidence intervals have better coverage accuracy, particularly when sample sizes are small. In the literature, generalized pivotal quantity-based inferences have been applied to many different problems. Gamage et al. [75] constructed a generalized confidence region for the difference between two mean vectors. Lee and Lin [76] developed confidence intervals for the ratio of the means of two normal populations. Tian and Wilding [77] presented a generalized variable approach for confidence interval estimation of a common correlation coefficient from several independent samples drawn from bivariate normal populations. Recently, Lai et al.[78] made use of a generalized approach to construct confidence intervals for the Youden index and its corresponding optimal cut-point. Further details on generalized confidence intervals can be found in Refs.[79] and [80].

The rest of the chapter is organized as follows. In Section 2, we introduce the induced-regression models for biomarker values from diseased and non-diseased populations. In Section 3, we derive GPQs for the AUC, and Youden index along with its cut-point. We also propose algorithms for computing the generalized confidence intervals for the AUC, and Youden index along with its cut-point. In section 4, simulation results are presented for evaluating coverage probabilities and the mean lengths of the GPQ-based intervals. We compare these probabilities and mean lengths with those of the normal approximation-based confidence intervals and bootstrap-based intervals. In Section 5, the proposed methods are applied to a real data set. Finally, Section 6 concludes the chapter with discussion.

3.2 AUC, Youden index and its associated cut-point in the presence of covariates

Two approaches have been used in the literature to model the relationship between the test/biomarker values and covariates. The first approach is to model the dependence of the ROC curve directly on the covariates. However this approach loses the connection with the cut-off value and does not allow the prediction of the sensitivity and specificity at a given cut-off conditional on covariates. The second approach is to directly model the covariate effects on the test results and through the modeling process obtain the covariate-adjusted ROC curve and its related summary measures. Faraggi [30] employed the second approach by using a simple linear regression model with normal error. In this session, we use the direct modeling approach and assume that the non-diseased test result (X) and diseased test result (Y) are linear functions of covariates ($\mathbf{Z_i}$):

$$X|\mathbf{Z_1} = \beta_1'\mathbf{Z_1} + \varepsilon_1 \tag{3.1}$$

$$Y|\mathbf{Z_2} = \beta_2'\mathbf{Z_2} + \varepsilon_2 \tag{3.2}$$

where $\mathbf{Z_i} = (Z_{i1}, Z_{i2}, \cdots, Z_{ip_i})'$, $i = 1, 2$, are p_i-dimensional covariates vectors associated with the non-diseased and diseased test results, respectively, $\mathbf{Z_1}$ and $\mathbf{Z_2}$ are assumed to have some common components, $\beta_i = (\beta_{i1}, \beta_{i2}, \cdots, \beta_{ip_i})'$ are p_i-dimensional column vectors of unknown parameters. p_i-dimensional column vectors of unknown parameters, and the error terms $\varepsilon_i \sim N(0, \sigma_i^2)$ and are independent random variables, and the error terms $\varepsilon_i \sim N(0, \sigma_i^2)$ and are independent random variables.

Under this model setting, at given covariates $\mathbf{Z_1} = \mathbf{z_1}$ and $\mathbf{Z_2} = \mathbf{z_2}$, the distribution of $X|\mathbf{Z_1}$ is the normal distribution with mean $\mu_{X|\mathbf{z_1}} \equiv E(X|\mathbf{Z_1} = \mathbf{z_1}) = \beta_1'\mathbf{z_1}$ and variance σ_1^2, and the distribution of $Y|\mathbf{Z_2}$ is the normal distribution with mean $\mu_{Y|\mathbf{z_2}} \equiv E(Y|\mathbf{Z_2} = \mathbf{z_2}) = \beta_2'\mathbf{z_2}$ and variance σ_2^2. From equations (3.1) and (3.2), we can derive the covariate-adjusted AUC as follows

$$A(\mathbf{z_1}, \mathbf{z_2}) = \text{Prob}(Y > X|\mathbf{Z_1} = \mathbf{z_1}, \mathbf{Z_2} = \mathbf{z_2}) = \Phi\{\delta(\mathbf{z_1}, \mathbf{z_2})\}, \tag{3.3}$$

where

$$\delta(\mathbf{z_1}, \mathbf{z_2}) = \frac{\mu_{Y|\mathbf{z_2}} - \mu_{X|\mathbf{z_1}}}{\sqrt{(\sigma_1^2 + \sigma_2^2)}}$$

with Φ being the standard normal cumulative distribution function.

The covariate-adjusted sensitivity and specificity at a given cut-point C are

$$q(\mathbf{z_1}, \mathbf{z_2}) = \Phi\left(\frac{\mu_{Y|\mathbf{z_2}} - C}{\sigma_2}\right),$$
$$p(\mathbf{z_1}, \mathbf{z_2}) = \Phi\left(\frac{C - \mu_{X|\mathbf{z_1}}}{\sigma_1}\right),$$

respectively.

The covariate-adjusted Youden index is defined as

$$YI(\mathbf{z_1}, \mathbf{z_2}) = \max_C\{p(\mathbf{z_1}, \mathbf{z_2}) + q(\mathbf{z_1}, \mathbf{z_2})\} - 1.$$

The covariate-adjusted optimal cut-off point $C^*(\mathbf{z_1}, \mathbf{z_2})$ is given by (see Ref. [81])

$$C^*(\mathbf{z_1}, \mathbf{z_2}) = \frac{\mu_{X|\mathbf{z_1}}(b^2 - 1) - a + b\sqrt{a^2 + (b^2 - 1)\sigma_1^2 \ln b^2}}{b^2 - 1}. \tag{3.4}$$

where $a = \mu_{Y|\mathbf{z_2}} - \mu_{X|\mathbf{z_1}}, b = \frac{\sigma_2}{\sigma_1}$.

Hence

$$YI(\mathbf{z_1}, \mathbf{z_2}) = \Phi\left(\frac{\mu_{Y|\mathbf{z_2}} - C^*(\mathbf{z_1}, \mathbf{z_2})}{\sigma_2}\right) + \Phi\left(\frac{C^*(\mathbf{z_1}, \mathbf{z_2}) - \mu_{X|\mathbf{z_1}}}{\sigma_1}\right) - 1. \tag{3.5}$$

If $\sigma_1 = \sigma_2$, $C^*(\mathbf{z_1}, \mathbf{z_2})$ is undefined and it can be replaced by

$$C^*(\mathbf{z_1}, \mathbf{z_2}) = \frac{\mu_{X|\mathbf{z_1}} + \mu_{Y|\mathbf{z_2}}}{2}, \tag{3.6}$$

which is the limit of (3.4) as $b \to 1$.

The covariate-adjusted AUC and Youden index along with its the optimal cut-off point defined as above are still unknown but they can be estimated by using the maximum likelihood estimates for β_i's and σ_i^2's. Let $\{(x_i, \mathbf{z}'_{i,1}) : i = 1, \cdots, m\}$ and $\{(y_j, \mathbf{z}'_{j,2}) : j = 1, \cdots, n\}$ be random samples of "non-diseased" subjects and "diseased" subjects from models (3.1)-(3.2) respectively, where $\mathbf{z}_{i,1} = (z_{i1,1}, z_{i2,1}, \cdots, z_{ip_1,1})'$ and $\mathbf{z}_{j,2} = (z_{j1,2}, z_{j2,2}, \cdots, z_{jp_2,2})'$ are the corresponding covariates values in the "non-diseased" and "diseased" samples. Our goal is to estimate $A(\mathbf{z_1}, \mathbf{z_2})$, $YI(\mathbf{z_1}, \mathbf{z_2})$ and $C^*(\mathbf{z_1}, \mathbf{z_2})$ at given $(\mathbf{z_1}, \mathbf{z_2})$ based on these samples. Let

$$\widetilde{Z}_1 = \begin{pmatrix} z_{11,1} & z_{12,1} & \cdots & z_{1p_1,1} \\ z_{21,1} & z_{22,1} & \cdots & z_{2p_1,1} \\ \vdots & \vdots & \vdots & \vdots \\ z_{m1,1} & z_{m2,1} & \cdots & z_{mp_1,1} \end{pmatrix}, \quad \widetilde{Z}_2 = \begin{pmatrix} z_{11,2} & z_{12,2} & \cdots & z_{1p_2,2} \\ z_{21,2} & z_{22,2} & \cdots & z_{2p_2,2} \\ \vdots & \vdots & \vdots & \vdots \\ z_{n1,2} & z_{m2,1} & \cdots & z_{mp_2,2} \end{pmatrix}.$$

Then, β_i's and σ_i^2's can be estimated by the following estimators based on the "non-diseased" and "diseased" samples, respectively, i.e.

$$\begin{aligned} \widehat{\beta}_1 &= (\widetilde{Z}_1'\widetilde{Z}_1)^{-1}\widetilde{Z}_1'\widetilde{X}, \\ \widehat{\beta}_2 &= (\widetilde{Z}_2'\widetilde{Z}_2)^{-1}\widetilde{Z}_2'\widetilde{Y}, \\ \widehat{\sigma}_1^2 &= (\widetilde{X}'\widetilde{X} - \widehat{\beta}_1\widetilde{Z}_1'\widetilde{X})/(m - p_1), \\ \widehat{\sigma}_1^2 &= (\widetilde{Y}'\widetilde{Y} - \widehat{\beta}_1\widetilde{Z}_1'\widetilde{Y})/(n - p_2), \end{aligned}$$

where $\widetilde{X} = (x_1, \cdots, x_m)'$ and $\widetilde{Y} = (y_1, \cdots, y_n)'$.

Substituting these estimates for the corresponding unknown parameters in (3.3)-(3.6), we obtain the point estimators $\widehat{A}(\mathbf{z_1}, \mathbf{z_2})$, $\widehat{YI}(\mathbf{z_1}, \mathbf{z_2})$ and $\widehat{C}^*(\mathbf{z_1}, \mathbf{z_2})$ for the covariate-adjusted

AUC and Youden index along with its optimal cut-off point. Using the asymptotic normality of $\widehat{\beta}_i$'s and $\widehat{\sigma}_i^2$'s and the delta method, it can be shown that these estimators are asymptotically normal (see Ref.[30]). Therefore, we can construct normal approximation-based confidence intervals for the covariate-adjusted AUC and Youden index along with its optimal cut-off point.

3.3 Generalized Confidence Intervals

The concept of the generalized confidence interval was introduced by Tsui and Weerahandi [74] and Weerahandi [67] . Suppose that Y is a random variable whose distribution depends on (θ, δ), where θ is a parameter of interest and δ is a nuisance parameter. Let y be the observed value of Y. $R(Y; y, \theta, \delta)$, a function of Y, y, θ, and δ, is called a generalized pivotal quantity (GPQ) if it satisfies the following two conditions:

1. $R(Y; y, \theta, \delta)$ has a distribution free of all unknown parameters.
2. The value of $R(Y; y, \theta, \delta)$ at $Y = y$ is θ, the parameter of interest.

Under model assumptions (3.1) and (3.2), $X|\mathbf{Z_1}$ and $Y|\mathbf{Z_2}$ are independent and follow normal distributions $N(\mu_{X|\mathbf{z_1}}, \sigma_1^2)$ and $N(\mu_{Y|\mathbf{z_2}}, \sigma_2^2)$, respectively. In the following, we will derive the GPQs of $\mu_{X|\mathbf{z_1}}$, $\mu_{Y|\mathbf{z_2}}$, σ_1^2, and σ_2^2 at given covariates $\mathbf{Z_1} = \mathbf{z_1}$, $\mathbf{Z_2} = \mathbf{z_2}$. Note that $\mu_{X|\mathbf{z_1}}$, $\mu_{Y|\mathbf{z_2}}$ can be consistently estimated by

$$\hat{\mu}_{X|\mathbf{z_1}} = \widehat{\beta}_1' \mathbf{z_1} \tag{3.7}$$

$$\hat{\mu}_{Y|\mathbf{z_2}} = \widehat{\beta}_2' \mathbf{z_2}. \tag{3.8}$$

Since $\hat{\mu}_{X|\mathbf{z_1}}$, and $\hat{\mu}_{Y|\mathbf{z_2}}$ are linear combinations of $\widehat{\beta}_1$ and $\widehat{\beta}_2$ which follow multivariate normal distributions, $\hat{\mu}_{X|\mathbf{z_1}}$ and $\hat{\mu}_{Y|\mathbf{z_2}}$ are also normally distributed. i.e.,

$$\hat{\mu}_{X|\mathbf{z_1}} \sim N\left(\mu_{X|\mathbf{z_1}}, \sigma_1^2 V_1\right)$$

$$\hat{\mu}_{Y|\mathbf{z_2}} \sim N\left(\mu_{Y|\mathbf{z_2}}, \sigma_2^2 V_2\right)$$

with

$$Var(\hat{\mu}_{X|\mathbf{z_1}}) = Var(\widehat{\beta}_1'\mathbf{z_1}) = \mathbf{z_1}'Var(\widehat{\beta}_1)\mathbf{z_1} = \sigma_1^2\mathbf{z_1}'(\widetilde{Z}_1'\widetilde{Z}_1)^{-1}\mathbf{z_1} \equiv \sigma_1^2 V_1 \tag{3.9}$$

$$Var(\hat{\mu}_{Y|\mathbf{z_2}}) = Var(\widehat{\beta}_2'\mathbf{z_2}) = \mathbf{z_2}'Var(\widehat{\beta}_2)\mathbf{z_2} = \sigma_2^2\mathbf{z_2}'(\widetilde{Z}_2'\widetilde{Z}_2)^{-1}\mathbf{z_2} \equiv \sigma_2^2 V_2 \tag{3.10}$$

Therefore, the GPQ for $\mu_{X|\mathbf{z_1}}$ is

$$\begin{aligned} R_{\mu_{X|\mathbf{z_1}}} &= \hat{\mu}_{X|\mathbf{z_1}} - \frac{\hat{\mu}_{X|\mathbf{z_1}} - \mu_{X|\mathbf{z_1}}}{\sigma_1\sqrt{V_1}} \times \sigma_1\sqrt{V_1}\frac{e_x}{e_X} \\ &= \hat{\mu}_{X|\mathbf{z_1}} - \frac{Z}{\sqrt{e_X^2/\sigma_1^2}} \times e_x\sqrt{V_1} \\ &= \hat{\mu}_{X|\mathbf{z_1}} - T_{m-p_1}\sqrt{\frac{m}{m-p_1}} \times e_x\sqrt{V_1}, \end{aligned} \tag{3.11}$$

where $e_X = \left\{\frac{\sum_i (X_i - \bar{X})^2}{m}\right\}^{1/2}$ with $\bar{X} = \sum_i X_i/m$, e_x is the observed value of e_X, and T_{m-p_1} is a chi-square random variable with degree of freedom $m - p_1$.

Similarly, the GPQ for $\mu_{Y|\mathbf{z_2}}$ is

$$R_{\mu_{Y|\mathbf{z_2}}} = \hat{\mu}_{Y|\mathbf{z_2}} - T_{n-p_2}\sqrt{\frac{n}{n-p_2}} \times e_y\sqrt{V_2}, \tag{3.12}$$

where $e_Y = \left\{\frac{\sum_j (Y_j - \bar{Y})^2}{n}\right\}^{1/2}$ with $\bar{Y} = \sum_j Y_j/n$, e_y is the observed value of e_Y, and T_{n-p_2} is a chi-square random variable with degree of freedom $n - p_2$.

The GPQs for σ_1^2 and σ_2^2 are:

$$R_{\sigma_1^2} = \frac{\sigma_1^2}{e_X^2} \times e_x^2 = \frac{me_x^2}{\chi^2_{m-p_1}} \tag{3.13}$$

$$R_{\sigma_2^2} = \frac{\sigma_2^2}{e_Y^2} \times e_y^2 = \frac{ne_y^2}{\chi^2_{n-p_2}} \tag{3.14}$$

Particularly, when $p_1 = p_2 = 2$, the above GPQ's are reduced to the GPQs given by [82].

Let $R_{\sigma_i} = \sqrt{R_{\sigma_i^2}}$, $i = 1, 2$, and

$$R_a = R_{\mu_{Y|\mathbf{z_2}}} - R_{\mu_{X|\mathbf{z_1}}}, \qquad R_b = \frac{R_{\sigma_2}}{R_{\sigma_1}}.$$

Obviously, R_{σ_i} is the GPQ for σ_i, R_a and R_b are the GPQs for a and b respectively.

Then by substituting $R_a, R_b, R_{\mu_{X|\mathbf{z_1}}}$ and $R_{\sigma_1^2}$ for the corresponding quantities $a, b, \mu_{X|\mathbf{z_1}}$ and σ_1^2 in (3.4) and (3.6), we get the GPQs for $C^*(\mathbf{z_1}, \mathbf{z_2})$:

$$R_{C^*} = \frac{R_{\mu_{X|\mathbf{z_1}}}(R_b^2 - 1) - R_a + R_b\sqrt{R_a^2 + (R_b^2 - 1)R_{\sigma_1^2} \ln R_b^2}}{R_b^2 - 1}. \tag{3.15}$$

When $\sigma_1^2 = \sigma_2^2$,

$$R_{C^*} = \frac{R_{\mu_{X|\mathbf{z_1}}} + R_{\mu_{Y|\mathbf{z_2}}}}{2}. \tag{3.16}$$

Similarly, the GPQs for $YI(\mathbf{z_1}, \mathbf{z_2})$ and $A(\mathbf{z_1}, \mathbf{z_2})$ are

$$R_{YI} = \Phi\left(\frac{R_{\mu_{Y|\mathbf{z_2}}} - R_{C^*}}{R_{\sigma_2}}\right) + \Phi\left(\frac{R_{C^*} - R_{\mu_{X|\mathbf{z_1}}}}{R_{\sigma_1}}\right), \tag{3.17}$$

and

$$R_{AUC} = \Phi\{R_{\delta(\mathbf{z_1},\mathbf{z_2})}\}, \tag{3.18}$$

respectively, where

$$R_{\delta(\mathbf{z_1},\mathbf{z_2})} = \frac{R_{\mu_{Y|\mathbf{z_2}}} - R_{\mu_{X|\mathbf{z_1}}}}{\sqrt{(R_{\sigma_1^2} + R_{\sigma_2^2})}}. \tag{3.19}$$

To construct $(1-\alpha)\%$ generalized confidence intervals for the covariate-adjusted AUC and Youden index along with its optimal cut-off point, we propose the following algorithm:

For the given "non-diseased" and "diseased" samples $\{(x_i, \mathbf{z}'_{i,1}) : i = 1, \cdots, m\}$ and $\{(y_j, \mathbf{z}'_{j,2}) : j = 1, \cdots, n\}$, and at given covariates $\mathbf{Z_1} = \mathbf{z_1}$ and $\mathbf{Z_2} = \mathbf{z_2}$,

1. Compute $e_x = \left\{\frac{\sum_i (x_i - \bar{x})^2}{m}\right\}^{1/2}$, $e_y = \left\{\frac{\sum_j (y_j - \bar{y})^2}{n}\right\}^{1/2}$, $\hat{\mu}_{X|\mathbf{z_1}}$ and $\hat{\mu}_{Y|\mathbf{z_2}}$ according to (3.7)-(3.8).

2. For $k = 1, \ldots, K$,

   - Generate T_{m-p_1} and T_{n-p_2} from Student's t-distribution with degrees of freedom $m - p_1$ and $n - p_2$ respectively;
   - Generate $\chi^2_{m-p_1}$ and $\chi^2_{n-p_2}$ from χ^2 distribution with degrees of freedom $m - p_1$ and $n - p_2$ respectively;
   - Compute $R_{\mu_{X|\mathbf{z_1}}}, R_{\mu_{Y|\mathbf{z_2}}}, R_{\sigma_x}$, and R_{σ_y} according to equations (3.11)-(3.14);
   - Compute $R_{C^*,k}$ following (3.15) or (3.16);
   - Compute $R_{AUC,k}$ and $R_{YI,k}$ following (3.18) and (3.17).

   (end k loop)

3. Compute the $100\alpha/2$-th percentile $R_{AUC,\alpha/2}$ and the $100(1-\alpha/2)$-th percentile $R_{AUC,(1-\alpha)/2}$ of $\{R_{AUC,1}, R_{AUC,2}, \ldots, R_{AUC,K}\}$. Then, $\left(R_{AUC,\alpha/2}, R_{AUC,(1-\alpha)/2}\right)$ is a $100(1-\alpha)\%$ generalized confidence interval for the covariate-adjusted AUC.

4. Compute the $100\alpha/2$-th percentile $R_{YI,\alpha/2}$ and the $100(1-\alpha/2)$-th percentile $R_{YI,(1-\alpha)/2}$ of $\{R_{YI,1}, R_{YI,2}, \ldots, R_{YI,K}\}$. Then, $\left(R_{YI,\alpha/2}, R_{YI,(1-\alpha)/2}\right)$ is a $100(1-\alpha)\%$ confidence interval for the covariate-adjusted YI.

5. Compute the $100\alpha/2$-th percentile $R_{C^*,\alpha/2}$ and the $100(1-\alpha/2)$-th percentile $R_{C^*,(1-\alpha)/2}$ of $\{R_{C^*,1}, R_{C^*,2}, \ldots, R_{C^*,K}\}$. Then, $\left(R_{C^*,\alpha/2}, R_{C^*,(1-\alpha)/2}\right)$ is a $100(1-\alpha)\%$ confidence interval for the covariate-adjusted optimal cut-off point.

3.4 Simulation Studies

In order to examine the finite sample performance of the generalized confidence intervals for the covariate-adjusted AUC and Youden index along with its optimal cut-point, we conduct extensive simulation studies to evaluate the coverage probabilities ("cp") and the average lengths ("al") of the confidence intervals. For comparison, we also provide coverage probabilities and average lengths of the bootstrap-based intervals for the covariate-adjusted AUC and Youden index along with its optimal cut-point. The coverage probabilities and average lengths of the normal approximation-based ("AN") intervals for the covariate-adjusted AUC are presented in the studies as well.

In the first simulation study, we choose $p_1 = p_2 = 2$, $\beta_{11} = 6$, $\beta_{12} = \beta_{22} = 1.5$, $\beta_{21} = 7.2$, and generate the "non-diseased" sample $\{(x_i, \mathbf{z}'_{i,1}) : i = 1, \cdots, m\}$ and the "diseased" sample $\{(y_j, \mathbf{z}'_{j,2}) : j = 1, \cdots, n\}$ from the following linear regression models:

Model 1:

$$X|\mathbf{Z_1} = 6 + 1.5Z_{12} + \varepsilon_1, \tag{3.20}$$

$$Y|\mathbf{Z_2} = 7.2 + 1.5Z_{22} + \varepsilon_2, \tag{3.21}$$

In the second simulation study, we choose $p_1 = p_2 = 3$. We keep the previous setting for the parameters above, and add $\beta_{13} = 1.8$, $\beta_{23} = 2$ to the models. The "non-diseased" and "diseased" samples are generated from the following models:

Model 2:

$$X|\mathbf{Z_1} = 6 + 1.5Z_{12} + 1.8Z_{13} + \varepsilon_1, \tag{3.22}$$

$$Y|\mathbf{Z_2} = 7.2 + 1.5Z_{22} + 2Z_{23} + \varepsilon_2. \tag{3.23}$$

In Model 1 and 2, $\mathbf{Z_1}$ and $\mathbf{Z_2}$ are covariates. The values of $\mathbf{Z_1}$ and $\mathbf{Z_2}$ among non-diseased and diseased groups are not necessarily the same. ε_i is generated from $N(0, \sigma_i^2)$ $(i = 1, 2)$, and ε_1 is independent of ε_2. Both Z_{12} and Z_{13} are generated from the uniform distribution on

$[1,5]$. Both Z_{22} and Z_{23} are generated from the uniform distribution on $[6,10]$. We choose $(m,n) = (10,10)$, $(30,30)$, $(20,50)$, $(50,20)$, $(50,50)$, and $(100,100)$ respectively. In the studies, the given covariate values are $\mathbf{z_1} = \mathbf{z_2} = (1, z_0)$ with z_0 being $2.0, 2.5, 3.0, 3.5, 4.0, 4.5$ respectively in Model (1). In Model (2), the given covariate values are $\mathbf{z_1} = \mathbf{z_2} = (1, z_1, z_2)$. We choose the following combinations of (z_1, z_2): $(2.0, 7.0)$, $(2.0, 9.5)$, $(2.5, 8.5)$, $(3.0, 8.0)$, $(3.5, 9.0)$, and $(4.5, 9.5)$.

Due to the complicated nature of the formulae for $\widehat{YI}(\mathbf{z_1}, \mathbf{z_2})$ and $\widehat{C}^*(\mathbf{z_1}, \mathbf{z_2})$, the derivation of their variances is quite complex. As Faraggi[30] suggested, we prefer to construct bootstrap-based confidence intervals for the covariate-adjusted AUC and Youden index along with its optimal cut-off point. We summarize the computation procedure of the proposed confidence intervals as follows:

1. Draw a bootstrap resample $\{(x_i^*, \mathbf{z}_{i,1}'^*) : i = 1, \cdots, m\}$ from the"non-diseased" sample $\{(x_i, \mathbf{z}_{i,1}') : i = 1, \cdots, m\}$, and a bootstrap resample $\{(y_j^*, \mathbf{z}_{j,2}'^*) : j = 1, \cdots, n\}$ form the "diseased" sample $\{(y_j, \mathbf{z}_{j,2}') : j = 1, \cdots, n\}$, respectively.

2. For $\hat{\theta} = \widehat{A}(\mathbf{z_1}, \mathbf{z_2})$, $\widehat{YI}(\mathbf{z_1}, \mathbf{z_2})$, $\widehat{C}^*(\mathbf{z_1}, \mathbf{z_2})$, compute the bootstrap copy θ^* of $\hat{\theta}$ from (3.3), (3.5) and (3.4), respectively.

3. Repeat the first two steps B times to obtain the bootstrap replications $\{\theta^{*b} : b = 1, 2, \cdots, B\}$. Then, the bootstrap estimator $V^*(\hat{\theta})$ for the variance of $\hat{\theta}$ is defined by

$$V^*(\hat{\theta}) = \frac{1}{B-1} \sum_{b=1}^{B} (\theta^{*b} - \overline{\theta}^*)^2$$

   where $\overline{\theta}^* = (1/B) \sum_{b=1}^{B} \theta^{*b}$.

4. Three $(1-\alpha)100\%$ $(0 < \alpha < 1)$ level bootstrap-based intervals for θ ($\theta = A(\mathbf{z_1}, \mathbf{z_2})$, $YI(\mathbf{z_1}, \mathbf{z_2})$, $C^*(\mathbf{z_1}, \mathbf{z_2})$) can be constructed as follows:

   The first interval, called BP interval, for θ is defined as

$$(\theta^{*([B\alpha/2])}, \theta^{*([B(1-\alpha/2)])}),$$

where $\theta^{*([B\alpha/2])}$ and $\theta^{*([B(1-\alpha/2)])}$ are the $\alpha/2$-th and $(1-\alpha/2)$-th quantiles of $\{\theta^{*b} : b = 1, 2, \cdots, B\}$, respectively.

The second interval, called BTI interval, for θ is defined as

$$(\widehat{\theta} - z_{1-\alpha/2}\sqrt{V^*(\hat{\theta})}, \widehat{\theta} + z_{1-\alpha/2}\sqrt{V^*(\hat{\theta})}),$$

where $\hat{\theta}$ is the estimate of θ from the original samples.

The third interval, called BTII interval, for θ is defined as

$$(\overline{\theta}^* - z_{1-\alpha/2}\sqrt{V^*(\hat{\theta})}, \overline{\theta}^* + z_{1-\alpha/2}\sqrt{V^*(\hat{\theta})}).$$

In the simulation studies, we generate $N = 1000$ "non-diseased" and "diseased" samples from Models 1 and 2, respectively, and choose $K = 1000$ for the calculation of the GPQ-based intervals. In the bootstrap procedure, we draw $B = 1000$ bootstrap samples from each of the "non-diseased" and "diseased" samples. Tables 1-3 display the coverage probabilities and the mean lengths of various intervals for the covariate-adjusted AUC and Youden index along with its optimal cut-point under **Model 1**. Tables 4-6 present the coverage probabilities and the mean lengths of the intervals under **Model 2**.

From Table 1-6, we can see that when sample sizes get bigger, the coverage probabilities of all the intervals are closer to the nominal level 95%, and the average lengths of the intervals become shorter. From Table 1, we observe that the coverage probabilities of the GPQ-based intervals are closer to the nominal level than those of the normal approximation-based AN intervals for all cases. Comparing the GPQ method with bootstrap-based methods, we can see that when sample sizes are small, the coverage probabilities of the bootstrap-based confidence intervals are far below the nominal level and the GPQ-based intervals still perform well.

In the multiple regression models (**Model 2**), the performances of the generalized confidence intervals are stable in every sample size and combination of given covariates' values. For AUC, we can see from Table 4 that the coverage probabilities of the bootstrap-based confidence intervals are far below the nominal level, and the average lengths of the bootstrap confidence intervals are longer than those of the generalized confidence intervals, especially when sample sizes are small. For Youden index, we can see from Table 5 that the coverage probabilities of the bootstrap-based interval are far below 95% and the average lengths are longer than those of the generalized confidence intervals when sample sizes are small ($n = m = 10$). In large sample size situations, although the average lengths of the bootstrap confidence intervals are shorter than those of the generalized confidence intervals, their coverage probabilities are below the nominal 95% level. For the optimal cut-point, we can see from Table 6 that in small sample size situations, the bootstrap-based intervals overcover the optimal cut-point and their average lengths are too big. Overall, the generalized confidence intervals outperform the other intervals in most cases considered here.

3.5 Real data analysis

For illustration of the proposed GPQ-based method, we present an application to a data set concerning diabetes diagnosis. This data set was first studied by Smith and Thompson [23] . It has also been analyzed in Refs. [30], [83] and [84]. The data come from a population-based pilot survey of diabetes mellitus in Cairo, Egypt, and consist of postprandial blood glucose measurements of 286 subjects obtained from a fingerstick. According to the gold standard criteria of the World Health Organization for diagnosing diabetes, 88 subjects were classified as diseased and 198 subjects as healthy. The age of the subject was considered as a relevant covariate in this example because glucose levels are expected to be higher for older people who do not suffer from diabetes (see Ref [23], for details).

To examine the effect of age in estimating the AUC and Youden index along with its optimal cut-off point, the following regression models are employed in the ROC analysis:

$$\begin{aligned} X|Z &= \beta_{11} + \beta_{12}Z + \varepsilon_1 \\ Y|Z &= \beta_{21} + \beta_{22}Z + \varepsilon_2, \end{aligned}$$

Table (3.1) The coverage probabilities and average lengths of the 95% confidence interval for the AUC in Model 1.

m	n	z_0	GPQ		AN		BP		BTI		$BTII$	
			cp	al	cp	al	cp	al	cp	al	cp	al
10	10	2.0	0.968	0.578	0.974	0.594	0.902	0.592	0.885	0.605	0.876	0.605
		2.5	0.966	0.506	0.976	0.532	0.916	0.528	0.902	0.532	0.884	0.532
		3.0	0.974	0.476	0.977	0.507	0.903	0.494	0.904	0.497	0.885	0.497
		3.5	0.969	0.505	0.980	0.534	0.910	0.523	0.903	0.528	0.882	0.528
		4.0	0.953	0.575	0.969	0.591	0.880	0.576	0.858	0.587	0.850	0.587
		4.5	0.963	0.661	0.978	0.667	0.902	0.658	0.853	0.692	0.859	0.692
30	30	2.0	0.951	0.336	0.967	0.361	0.944	0.333	0.929	0.335	0.930	0.335
		2.5	0.956	0.282	0.975	0.315	0.927	0.283	0.922	0.284	0.915	0.284
		3.0	0.950	0.262	0.965	0.298	0.937	0.261	0.939	0.262	0.927	0.262
		3.5	0.952	0.282	0.964	0.314	0.941	0.281	0.942	0.281	0.938	0.281
		4.0	0.955	0.335	0.968	0.359	0.918	0.331	0.911	0.332	0.908	0.332
		4.5	0.950	0.403	0.967	0.420	0.937	0.397	0.927	0.401	0.929	0.401
20	50	2.0	0.951	0.297	0.969	0.315	0.925	0.294	0.918	0.296	0.919	0.296
		2.5	0.959	0.250	0.977	0.274	0.932	0.248	0.930	0.248	0.928	0.248
		3.0	0.947	0.232	0.970	0.258	0.941	0.229	0.943	0.230	0.937	0.230
		3.5	0.944	0.251	0.961	0.274	0.941	0.248	0.941	0.248	0.934	0.248
		4.0	0.958	0.297	0.968	0.315	0.915	0.295	0.910	0.297	0.908	0.297
		4.5	0.940	0.358	0.961	0.370	0.947	0.357	0.927	0.360	0.931	0.360
50	20	2.0	0.948	0.387	0.958	0.413	0.904	0.378	0.893	0.380	0.886	0.380
		2.5	0.951	0.332	0.962	0.365	0.922	0.323	0.915	0.323	0.906	0.323
		3.0	0.944	0.305	0.964	0.344	0.908	0.300	0.906	0.301	0.894	0.301
		3.5	0.949	0.328	0.965	0.363	0.903	0.325	0.909	0.326	0.890	0.326
		4.0	0.948	0.384	0.967	0.410	0.922	0.381	0.910	0.383	0.906	0.383
		4.5	0.949	0.456	0.968	0.473	0.909	0.450	0.881	0.455	0.879	0.455
50	50	2.0	0.962	0.260	0.983	0.282	0.956	0.256	0.953	0.257	0.945	0.257
		2.5	0.952	0.219	0.970	0.245	0.936	0.215	0.937	0.215	0.934	0.215
		3.0	0.955	0.202	0.977	0.232	0.933	0.201	0.936	0.202	0.927	0.202
		3.5	0.957	0.218	0.973	0.245	0.917	0.216	0.917	0.216	0.917	0.216
		4.0	0.944	0.259	0.967	0.281	0.932	0.258	0.928	0.259	0.925	0.259
		4.5	0.959	0.312	0.972	0.329	0.948	0.311	0.940	0.313	0.939	0.313
100	100	2.0	0.943	0.183	0.961	0.200	0.953	0.183	0.949	0.183	0.947	0.183
		2.5	0.953	0.153	0.974	0.174	0.940	0.153	0.940	0.154	0.939	0.154
		3.0	0.956	0.142	0.980	0.165	0.940	0.143	0.943	0.143	0.937	0.143
		3.5	0.937	0.154	0.969	0.174	0.936	0.153	0.940	0.154	0.935	0.154
		4.0	0.939	0.185	0.963	0.201	0.941	0.183	0.937	0.183	0.936	0.183
		4.5	0.950	0.224	0.964	0.237	0.947	0.223	0.935	0.223	0.934	0.223

Table (3.2) The coverage probabilities and average lengths of the 95% confidence interval for the Youden Index in Model 1.

m	n	z_0	*GPQ*		*BP*		*BTI*		*BTII*	
			cp	al	cp	al	cp	al	cp	al
10	10	2.0	0.962	0.685	0.912	0.720	0.871	0.749	0.871	0.749
		2.5	0.964	0.623	0.914	0.681	0.907	0.691	0.891	0.691
		3.0	0.968	0.598	0.901	0.656	0.906	0.660	0.885	0.660
		3.5	0.954	0.625	0.912	0.678	0.908	0.687	0.888	0.687
		4.0	0.951	0.683	0.879	0.709	0.848	0.735	0.841	0.735
		4.5	0.965	0.752	0.904	0.763	0.826	0.821	0.841	0.821
30	30	2.0	0.950	0.432	0.940	0.437	0.936	0.438	0.925	0.438
		2.5	0.946	0.372	0.924	0.377	0.921	0.377	0.914	0.377
		3.0	0.945	0.348	0.928	0.353	0.939	0.352	0.927	0.352
		3.5	0.949	0.371	0.936	0.376	0.938	0.375	0.933	0.375
		4.0	0.958	0.431	0.925	0.436	0.918	0.437	0.914	0.437
		4.5	0.951	0.501	0.935	0.504	0.904	0.510	0.912	0.510
20	50	2.0	0.942	0.398	0.920	0.395	0.915	0.397	0.912	0.397
		2.5	0.946	0.346	0.919	0.341	0.922	0.342	0.918	0.342
		3.0	0.950	0.324	0.929	0.321	0.944	0.321	0.931	0.321
		3.5	0.931	0.348	0.931	0.340	0.929	0.340	0.919	0.340
		4.0	0.957	0.398	0.924	0.395	0.917	0.397	0.909	0.397
		4.5	0.947	0.465	0.933	0.464	0.912	0.469	0.912	0.469
50	20	2.0	0.956	0.481	0.908	0.485	0.890	0.487	0.886	0.487
		2.5	0.953	0.419	0.922	0.429	0.930	0.426	0.919	0.426
		3.0	0.945	0.388	0.919	0.404	0.923	0.401	0.915	0.401
		3.5	0.952	0.417	0.914	0.429	0.909	0.427	0.899	0.427
		4.0	0.951	0.479	0.929	0.496	0.914	0.497	0.905	0.497
		4.5	0.945	0.555	0.908	0.558	0.882	0.569	0.881	0.569
50	50	2.0	0.969	0.338	0.952	0.339	0.947	0.339	0.940	0.339
		2.5	0.954	0.287	0.931	0.287	0.937	0.286	0.931	0.286
		3.0	0.942	0.267	0.937	0.269	0.940	0.269	0.931	0.269
		3.5	0.953	0.287	0.930	0.287	0.934	0.287	0.925	0.287
		4.0	0.947	0.337	0.936	0.338	0.926	0.338	0.924	0.338
		4.5	0.959	0.400	0.940	0.404	0.934	0.406	0.929	0.406
100	100	2.0	0.937	0.240	0.946	0.240	0.949	0.240	0.946	0.240
		2.5	0.953	0.202	0.936	0.202	0.942	0.203	0.938	0.203
		3.0	0.947	0.188	0.940	0.188	0.945	0.188	0.939	0.188
		3.5	0.939	0.202	0.942	0.203	0.949	0.203	0.942	0.203
		4.0	0.940	0.240	0.934	0.240	0.935	0.240	0.933	0.240
		4.5	0.949	0.289	0.946	0.289	0.944	0.290	0.940	0.290

Table (3.3) The coverage probabilities and average lengths of the 95% confidence interval for the Cut-point in Model 1.

m	n	z_0	GPQ		BP		BTI		$BTII$	
			cp	al	cp	al	cp	al	cp	al
10	10	2.0	0.971	5.273	0.931	5.352	0.965	13.467	0.963	13.467
		2.5	0.958	3.757	0.908	3.490	0.965	9.146	0.953	9.146
		3.0	0.959	3.154	0.907	2.870	0.956	7.687	0.944	7.687
		3.5	0.963	3.580	0.910	3.154	0.966	8.736	0.949	8.736
		4.0	0.971	5.007	0.925	4.889	0.963	12.861	0.964	12.861
		4.5	0.971	6.909	0.945	8.204	0.980	22.096	0.975	22.096
30	30	2.0	0.965	1.278	0.936	1.158	0.947	1.398	0.941	1.398
		2.5	0.961	1.038	0.915	0.941	0.924	1.008	0.918	1.008
		3.0	0.959	0.955	0.918	0.874	0.941	0.934	0.925	0.934
		3.5	0.951	1.038	0.928	0.930	0.940	0.997	0.936	0.997
		4.0	0.951	1.260	0.925	1.164	0.941	1.375	0.932	1.375
		4.5	0.948	1.664	0.938	1.599	0.948	2.244	0.947	2.244
20	50	2.0	0.949	1.469	0.911	1.304	0.925	1.354	0.918	1.354
		2.5	0.954	1.202	0.904	1.062	0.920	1.097	0.908	1.097
		3.0	0.948	1.113	0.905	0.986	0.917	1.011	0.912	1.011
		3.5	0.940	1.229	0.920	1.074	0.938	1.121	0.930	1.121
		4.0	0.945	1.452	0.909	1.295	0.932	1.391	0.920	1.391
		4.5	0.950	1.937	0.916	1.646	0.934	1.845	0.926	1.845
50	20	2.0	0.942	1.169	0.941	1.227	0.962	2.248	0.959	2.248
		2.5	0.948	0.977	0.933	0.961	0.957	1.437	0.942	1.437
		3.0	0.958	0.883	0.927	0.859	0.939	1.091	0.934	1.091
		3.5	0.965	0.962	0.930	0.995	0.943	1.610	0.940	1.610
		4.0	0.957	1.168	0.942	1.215	0.952	2.054	0.951	2.054
		4.5	0.959	1.572	0.956	1.825	0.971	3.933	0.970	3.933
50	50	2.0	0.955	0.881	0.926	0.819	0.937	0.832	0.931	0.832
		2.5	0.948	0.755	0.944	0.703	0.951	0.705	0.946	0.705
		3.0	0.946	0.697	0.931	0.660	0.935	0.660	0.932	0.660
		3.5	0.942	0.741	0.923	0.703	0.933	0.708	0.923	0.708
		4.0	0.949	0.887	0.935	0.842	0.948	0.849	0.944	0.849
		4.5	0.939	1.075	0.939	1.027	0.947	1.050	0.944	1.050
100	100	2.0	0.939	0.593	0.942	0.579	0.942	0.579	0.942	0.579
		2.5	0.961	0.510	0.947	0.496	0.951	0.496	0.944	0.496
		3.0	0.944	0.476	0.943	0.465	0.945	0.465	0.945	0.465
		3.5	0.954	0.510	0.948	0.496	0.949	0.496	0.944	0.496
		4.0	0.950	0.597	0.947	0.581	0.947	0.580	0.946	0.580
		4.5	0.945	0.722	0.946	0.699	0.953	0.698	0.949	0.698

Table (3.4) The coverage probabilities and average lengths of the 95% confidence interval for the AUC in Model 2.

m	n	(z_1, z_2)	*GPQ*		*AN*		*BP*		*BTI*		*BTII*	
			cp	al	cp	al	cp	al	cp	al	cp	al
10	10	(2.0,7.0)	0.953	0.402	0.949	0.369	0.900	0.555	0.911	0.602	0.907	0.602
		(2.0,9.5)	0.952	0.274	0.955	0.265	0.907	0.554	0.939	0.618	0.941	0.618
		(2.5,8.5)	0.962	0.192	0.976	0.209	0.860	0.370	0.897	0.417	0.882	0.417
		(3.0,8.0)	0.962	0.173	0.983	0.197	0.866	0.334	0.897	0.378	0.876	0.378
		(3.5,9.0)	0.966	0.213	0.972	0.221	0.869	0.434	0.923	0.489	0.918	0.489
		(4.5,9.5)	0.966	0.306	0.961	0.288	0.911	0.616	0.944	0.694	0.944	0.694
30	30	(2.0,7.0)	0.956	0.380	0.972	0.419	0.935	0.252	0.910	0.258	0.905	0.258
		(2.0,9.5)	0.948	0.252	0.978	0.314	0.930	0.227	0.897	0.235	0.896	0.235
		(2.5,8.5)	0.946	0.187	0.990	0.270	0.908	0.165	0.892	0.168	0.878	0.168
		(3.0,8.0)	0.958	0.170	0.995	0.259	0.910	0.152	0.901	0.155	0.886	0.155
		(3.5,9.0)	0.949	0.203	0.981	0.280	0.913	0.180	0.885	0.184	0.883	0.184
		(4.5,9.5)	0.955	0.289	0.972	0.345	0.921	0.264	0.895	0.274	0.901	0.274
20	50	(2.0,7.0)	0.963	0.341	0.974	0.365	0.940	0.220	0.919	0.224	0.915	0.224
		(2.0,9.5)	0.959	0.225	0.985	0.268	0.916	0.205	0.889	0.211	0.889	0.211
		(2.5,8.5)	0.959	0.158	0.993	0.219	0.927	0.146	0.911	0.147	0.894	0.147
		(3.0,8.0)	0.966	0.148	0.995	0.215	0.906	0.137	0.899	0.138	0.887	0.138
		(3.5,9.0)	0.947	0.178	0.980	0.233	0.925	0.163	0.903	0.166	0.899	0.166
		(4.5,9.5)	0.952	0.248	0.977	0.287	0.942	0.235	0.909	0.242	0.915	0.242
50	20	(2.0,7.0)	0.947	0.453	0.962	0.492	0.891	0.295	0.871	0.305	0.866	0.305
		(2.0,9.5)	0.949	0.307	0.972	0.385	0.902	0.274	0.859	0.288	0.859	0.288
		(2.5,8.5)	0.965	0.219	0.990	0.321	0.904	0.191	0.878	0.197	0.860	0.197
		(3.0,8.0)	0.956	0.205	0.992	0.313	0.888	0.175	0.878	0.180	0.857	0.180
		(3.5,9.0)	0.955	0.246	0.982	0.340	0.887	0.211	0.867	0.219	0.854	0.219
		(4.5,9.5)	0.946	0.346	0.968	0.414	0.915	0.322	0.878	0.340	0.886	0.340
50	50	(2.0,7.0)	0.942	0.294	0.968	0.326	0.924	0.194	0.916	0.196	0.916	0.196
		(2.0,9.5)	0.937	0.188	0.968	0.238	0.933	0.174	0.909	0.178	0.910	0.178
		(2.5,8.5)	0.962	0.138	0.997	0.203	0.936	0.129	0.923	0.130	0.909	0.130
		(3.0,8.0)	0.948	0.128	0.996	0.197	0.917	0.120	0.914	0.121	0.901	0.121
		(3.5,9.0)	0.954	0.151	0.993	0.210	0.931	0.141	0.917	0.143	0.909	0.143
		(4.5,9.5)	0.958	0.207	0.983	0.251	0.925	0.197	0.903	0.202	0.912	0.202
100	100	(2.0,7.0)	0.962	0.397	0.950	0.365	0.941	0.137	0.937	0.138	0.935	0.138
		(2.0,9.5)	0.957	0.272	0.964	0.263	0.935	0.120	0.917	0.121	0.917	0.121
		(2.5,8.5)	0.968	0.192	0.983	0.208	0.949	0.092	0.938	0.093	0.931	0.093
		(3.0,8.0)	0.973	0.176	0.988	0.199	0.937	0.086	0.932	0.087	0.924	0.087
		(3.5,9.0)	0.960	0.210	0.975	0.219	0.930	0.100	0.911	0.101	0.908	0.101
		(4.5,9.5)	0.952	0.301	0.948	0.283	0.945	0.136	0.924	0.138	0.924	0.138

Table (3.5) The coverage probabilities and average lengths of the 95% confidence interval for the Youden Index in Model 2.

m	n	(z_1, z_2)	GPQ		BP		BTI		$BTII$	
			cp	al	cp	al	cp	al	cp	al
10	10	(2.0,7.0)	0.950	0.619	0.893	0.756	0.871	0.817	0.849	0.817
		(2.0,9.5)	0.948	0.508	0.900	0.745	0.871	0.826	0.870	0.826
		(2.5,8.5)	0.954	0.390	0.846	0.605	0.858	0.638	0.818	0.638
		(3.0,8.0)	0.963	0.358	0.846	0.582	0.892	0.608	0.836	0.608
		(3.5,9.0)	0.962	0.423	0.865	0.654	0.865	0.705	0.832	0.705
		(4.5,9.5)	0.967	0.549	0.907	0.784	0.858	0.883	0.864	0.883
30	30	(2.0,7.0)	0.949	0.584	0.931	0.441	0.922	0.443	0.911	0.443
		(2.0,9.5)	0.949	0.455	0.921	0.436	0.906	0.442	0.897	0.442
		(2.5,8.5)	0.947	0.334	0.907	0.322	0.915	0.324	0.891	0.324
		(3.0,8.0)	0.952	0.301	0.899	0.294	0.930	0.295	0.890	0.295
		(3.5,9.0)	0.949	0.370	0.911	0.355	0.901	0.358	0.881	0.358
		(4.5,9.5)	0.957	0.504	0.924	0.485	0.887	0.494	0.885	0.494
20	50	(2.0,7.0)	0.967	0.541	0.914	0.393	0.911	0.395	0.892	0.395
		(2.0,9.5)	0.959	0.421	0.908	0.393	0.887	0.397	0.876	0.397
		(2.5,8.5)	0.964	0.308	0.899	0.291	0.917	0.293	0.889	0.293
		(3.0,8.0)	0.967	0.285	0.883	0.268	0.893	0.269	0.876	0.269
		(3.5,9.0)	0.944	0.344	0.906	0.322	0.906	0.325	0.888	0.325
		(4.5,9.5)	0.956	0.459	0.935	0.436	0.900	0.443	0.896	0.443
50	20	(2.0,7.0)	0.944	0.654	0.895	0.509	0.887	0.512	0.876	0.512
		(2.0,9.5)	0.951	0.521	0.909	0.509	0.875	0.517	0.864	0.517
		(2.5,8.5)	0.961	0.378	0.906	0.374	0.923	0.376	0.885	0.376
		(3.0,8.0)	0.950	0.342	0.891	0.339	0.910	0.340	0.881	0.340
		(3.5,9.0)	0.950	0.425	0.895	0.412	0.899	0.415	0.878	0.415
		(4.5,9.5)	0.944	0.572	0.910	0.562	0.868	0.575	0.866	0.575
50	50	(2.0,7.0)	0.943	0.466	0.925	0.341	0.918	0.342	0.915	0.342
		(2.0,9.5)	0.939	0.352	0.935	0.342	0.926	0.344	0.917	0.344
		(2.5,8.5)	0.959	0.254	0.941	0.248	0.939	0.249	0.929	0.249
		(3.0,8.0)	0.943	0.229	0.921	0.225	0.930	0.226	0.913	0.226
		(3.5,9.0)	0.954	0.284	0.929	0.277	0.931	0.278	0.918	0.278
		(4.5,9.5)	0.945	0.387	0.926	0.380	0.901	0.383	0.903	0.383
100	100	(2.0,7.0)	0.956	0.615	0.938	0.241	0.936	0.241	0.931	0.241
		(2.0,9.5)	0.949	0.505	0.934	0.242	0.926	0.243	0.920	0.243
		(2.5,8.5)	0.968	0.391	0.941	0.176	0.941	0.176	0.933	0.176
		(3.0,8.0)	0.966	0.361	0.928	0.160	0.937	0.160	0.927	0.160
		(3.5,9.0)	0.957	0.421	0.928	0.195	0.928	0.196	0.914	0.196
		(4.5,9.5)	0.950	0.539	0.943	0.270	0.932	0.272	0.928	0.272

Table (3.6) The coverage probabilities and average lengths of the 95% confidence interval for the Cut-point in Model 2.

m	n	(z_1, z_2)	GPQ		BP		BTI		$BTII$	
			cp	al	cp	al	cp	al	cp	al
10	10	(2.0,7.0)	0.953	3.871	0.970	4.969	0.996	17.032	0.994	17.032
		(2.0,9.5)	0.947	2.655	0.975	5.847	0.996	22.925	0.995	22.925
		(2.5,8.5)	0.954	1.645	0.969	3.048	0.993	10.243	0.984	10.243
		(3.0,8.0)	0.941	1.422	0.966	2.362	0.995	7.008	0.990	7.008
		(3.5,9.0)	0.948	1.918	0.965	3.686	0.988	13.441	0.983	13.441
		(4.5,9.5)	0.948	3.112	0.969	7.918	0.988	29.549	0.985	29.549
30	30	(2.0,7.0)	0.952	1.725	0.931	1.131	0.938	1.160	0.934	1.160
		(2.0,9.5)	0.957	1.364	0.939	1.310	0.940	1.320	0.939	1.320
		(2.5,8.5)	0.946	0.887	0.928	0.852	0.944	0.850	0.935	0.850
		(3.0,8.0)	0.955	0.772	0.927	0.743	0.943	0.742	0.933	0.742
		(3.5,9.0)	0.955	1.037	0.924	1.007	0.933	1.014	0.925	1.014
		(4.5,9.5)	0.953	1.564	0.938	1.512	0.938	1.577	0.937	1.577
20	50	(2.0,7.0)	0.954	2.042	0.923	1.337	0.945	1.335	0.932	1.335
		(2.0,9.5)	0.946	1.568	0.905	1.541	0.923	1.535	0.916	1.535
		(2.5,8.5)	0.963	1.021	0.907	1.006	0.940	1.004	0.920	1.004
		(3.0,8.0)	0.950	0.898	0.890	0.875	0.935	0.874	0.903	0.874
		(3.5,9.0)	0.947	1.213	0.912	1.167	0.935	1.163	0.927	1.163
		(4.5,9.5)	0.948	1.810	0.922	1.727	0.936	1.731	0.927	1.731
50	20	(2.0,7.0)	0.951	1.699	0.953	1.132	0.960	1.572	0.959	1.572
		(2.0,9.5)	0.954	1.273	0.943	1.321	0.961	1.851	0.957	1.851
		(2.5,8.5)	0.946	0.816	0.950	0.834	0.955	0.894	0.952	0.894
		(3.0,8.0)	0.958	0.711	0.959	0.715	0.965	0.719	0.963	0.719
		(3.5,9.0)	0.952	0.949	0.954	0.973	0.964	1.041	0.961	1.041
		(4.5,9.5)	0.966	1.476	0.947	1.588	0.962	2.587	0.956	2.587
50	50	(2.0,7.0)	0.954	1.211	0.930	0.843	0.932	0.843	0.931	0.843
		(2.0,9.5)	0.945	1.010	0.937	0.981	0.934	0.980	0.935	0.980
		(2.5,8.5)	0.949	0.662	0.927	0.641	0.936	0.640	0.933	0.640
		(3.0,8.0)	0.953	0.575	0.927	0.561	0.939	0.561	0.929	0.561
		(3.5,9.0)	0.943	0.771	0.943	0.747	0.947	0.746	0.942	0.746
		(4.5,9.5)	0.944	1.132	0.940	1.110	0.941	1.107	0.938	1.107
100	100	(2.0,7.0)	0.951	3.713	0.934	0.585	0.944	0.584	0.934	0.584
		(2.0,9.5)	0.947	2.606	0.939	0.681	0.941	0.681	0.936	0.681
		(2.5,8.5)	0.950	1.629	0.932	0.447	0.929	0.448	0.936	0.448
		(3.0,8.0)	0.953	1.423	0.943	0.391	0.951	0.391	0.944	0.391
		(3.5,9.0)	0.953	1.943	0.956	0.518	0.957	0.517	0.952	0.517
		(4.5,9.5)	0.948	2.989	0.944	0.766	0.945	0.765	0.944	0.765

where X and Y are the transformed (postprandial blood glucose) biomarker values (which is $-(\text{biomarker value})^{-\frac{1}{2}}$ as Farragi [30] suggested) for the non-diseased and diseased subjects, respectively, and Z denotes the age of the subjects. The transformed biomarker values comply with the normality assumption for the underlying models.

Based on our simulation results, we construct generalized confidence intervals for the age-adjusted AUC and Youden index along with its optimal cut-off point at given Z (age). Figures 1-3 provide estimates for the area under the ROC curve and Youden index along with its the optimal cut-off point as a function of age and the corresponding pointwise 95% level generalized confidence intervals. In Figure 3, the values of the cut-off points have been transformed back to the original biomarker values. From Figures 1-2, we can see that the diagnostic accuracy of the biomarker for individuals with age < 50 years is high (AUC > 0.90, YI > 0.7). From Figure 3, we observe that the estimated cut-off value increases as individuals get older. Using a fixed cut-off value regardless of age would be misleading. Additionally, we observe that the widths of all three types of intervals for people with ages between 40 and 60 are much shorter than those for people with ages < 40 or age > 60. Thus inferences on the diagnostic accuracy of the biomarker for people with ages between 40 and 60 will be more precise. These results indicate substantial effects of age on the estimation of AUC and Youden index along with its optimal cut-off point in the diagnosis of diabetes by using postprandial blood glucose measurements. This conclusion is consistent with Faraggi [30] .

3.6 Discussion

Covariates are important in the evaluation of the diagnostic accuracy of a biomarker/medical test. Ignoring the covariates' effects may lead to biased estimation of the diagnostic accuracy and even wrong conclusions. Pepe [7] gave an introduction to why and how to adjust for covariates in ROC analysis. Pardo-Fernandez et al. [85] gave an excellent review on ROC curve analysis in the presence of covariates. One important approach to incorporate covariates to the ROC analysis is through regression models. In a parametric

Figure (3.1) AUC curve as a function of age with 95% confidence intervals based on GPQ method

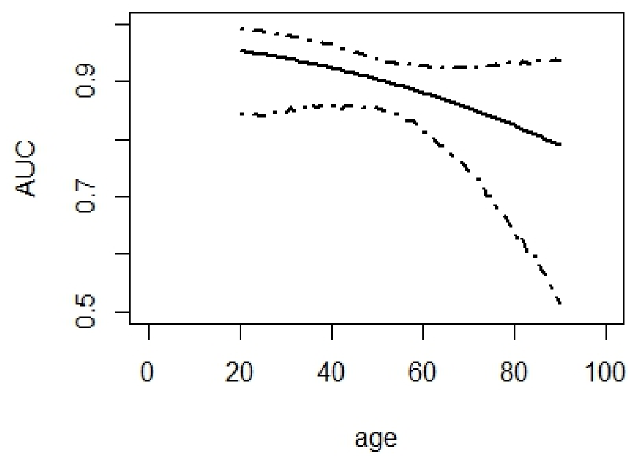


Figure (3.2) YI curve as a function of age with 95% confidence intervals based on GPQ method

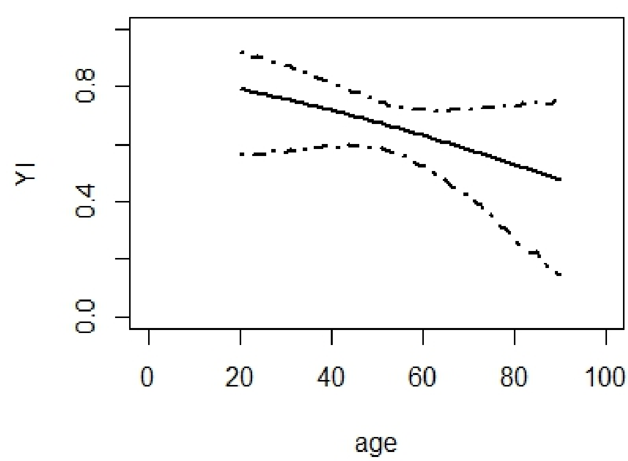
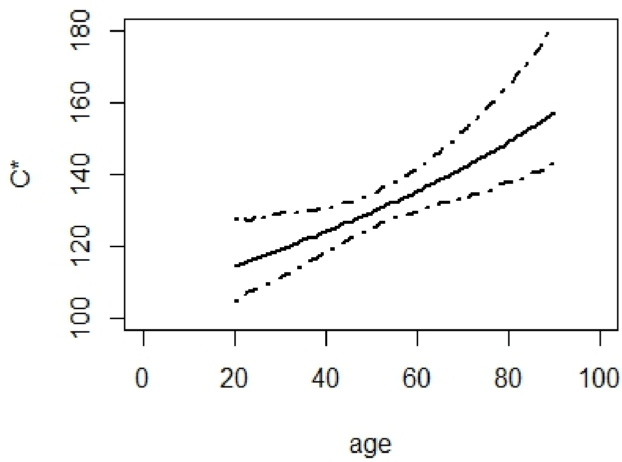


Figure (3.3) C* curve as a function of age with 95% confidence intervals based on GPQ method



framework, Faraggi [30] used simple linear regression models for the conditional means with normal errors, in both non-diseased and diseased populations, and provided a simple method for inferences on covariate-adjusted ROC curve. It is well known that bi-normal models play an important role in parametric ROC curve analysis, and the GPQ-based methods can provide "exact" interval estimation for AUC and Youden index under bi-normal models for test results (see Refs.[86] and [87]). In this paper, we have proposed GPQ-based intervals for covariate-adjusted AUC and Youden index along with its optimal cut-off point. Our simulation results have shown that the proposed methods outperform existing parametric methods under the same parametric linear models setting, particularly for small to moderate sized samples which are more applicable and practical in second or third phase medical diagnostic trial studies. As Faraggi [30] indicated, although the method is limited by the normality assumption, it can be extended to many non-normal situations by using Box-Cox-type transformations. Further research will focus on non-linear/non-parametric regression modeling for test results in ROC analysis when covariates are present.

Chapter 4

UNIFORM TEST FOR PREDICTIVE REGRESSION WITH AR ERRORS

This chapter is based on the accepted paper by C. Li, D. Li and L. Peng, *Uniform Test for Predictive Regression with AR Errors*, *Journal of Business & Economic Statistics* 11 Jun 2015.

4.1 Review and Outline

Referring to what we stated in Chapter 1 1.2, in this chapter, we investigate the possibility of extending the unified approach in Zhu, Cai and Peng (2014 [1]) to the case when $\{U_t\}$ follows an AR(p) process. When $\{U_t\}$ is an α-mixing sequence, estimation and test are proposed by Cai and Wang (2014 [64]), which do not lead to a unified procedure. We organize this chapter as follows. Section 4.2 presents methodologies and theoretical results. A simulation study and a real data analysis are given in Sections 4.3 and 4.4, respectively. All proofs are put in Appendix A.

4.2 Methodologies and Theoretical Results

4.2.1 Profile Empirical Likelihood

To better appreciate the proposed methodologies, we start with a simpler predictive regression model without an intercept. That is, we assume the observations $\{(X_t, Y_t)\}_{t=1}^n$ follow from

$$Y_t = \beta X_{t-1} + U_t, \quad X_t = \theta + \phi X_{t-1} + e_t, \quad B(L)e_t = V_t, \quad \varepsilon_t = U_t + \sum_{j=1}^{p} \gamma_j U_{t-j}, \qquad (4.1)$$

where $(\varepsilon_1, V_1), ..., (\varepsilon_n, V_n)$ are independent and identically distributed random vectors. Obviously the least squares estimator $\hat{\beta}_{LS} = \sum_{t=1}^n Y_t X_{t-1} / \sum_{t=1}^n X_{t-1}^2$ is inefficient. By taking the structure of U_t into account, one may consider to minimize the following least squares

$$\sum_{t=1}^{n}\Big\{Y_t-\beta X_{t-1}+\sum_{j=1}^{p}\gamma_j\big(Y_{t-j}-\beta X_{t-j-1}\big)\Big\}^2,$$

which implies that the proposed new least squares estimator solves the following score equations

$$\sum_{t=1}^{n}\Big\{Y_t-\beta X_{t-1}+\sum_{j=1}^{p}\gamma_j(Y_{t-j}-\beta X_{t-j-1})\Big\}(Y_{t-j}-\beta X_{t-j-1})=0,\quad \text{for } j=1,2,...,p \tag{4.2}$$

and

$$\sum_{t=1}^{n}\Big\{Y_t-\beta X_{t-1}+\sum_{j=1}^{p}\gamma_j(Y_{t-j}-\beta X_{t-j-1})\Big\}\big(X_{t-1}+\sum_{j=1}^{p}\gamma_j X_{t-j-1}\big)=0. \tag{4.3}$$

This simple idea of taking the error structure into account has appeared in the literature; see Xiao, Linton, Carroll, Mammen (2003 [88]) and Liu, Chen and Yao (2010 [89]) for nonparametric regression models; Hall and Yao (2003 [90]) for parametric regression models; Hill, Li and Peng (2014 [91]) for an AR process with a possible near unit root.

In order to construct an interval for β without estimating the asymptotic variance, one may apply the profile empirical likelihood method to the above equations as in Qin and Lawless (1994 [92]). However, when $\{X_t\}$ is nearly integrated, Wilks theorem fails for the above profile empirical likelihood method. Like the uniform estimation in Zhu, Cai and Peng (2014), we propose to apply the profile empirical likelihood method to equations (4.2) and the following weighted version of (4.3)

$$\sum_{t=1}^{n}\left\{Y_t-\beta X_{t-1}+\sum_{j=1}^{p}\gamma_j(Y_{t-j}-\beta X_{t-j-1})\right\}\left\{\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}}+\sum_{j=1}^{p}\gamma_j\frac{X_{t-j-1}}{\sqrt{1+X_{t-j-1}^2}}\right\}=0. \tag{4.4}$$

More specifically, write $Z_t(\beta,\gamma) := (Z_{t,1}(\beta,\gamma),\cdots,Z_{t,p+1}(\beta,\gamma))^T \in \mathbb{R}^{p+1}$, where

$$Z_{t,j}(\beta,\gamma) = \left\{Y_t - \beta X_{t-1} + \sum_{k=1}^{p}\gamma_k(Y_{t-k} - \beta X_{t-k-1})\right\}(Y_{t-j} - \beta X_{t-j-1}) \text{ for } j = 1,\cdots,p,$$

and

$$Z_{t,p+1}(\beta,\gamma) = \left\{Y_t - \beta X_{t-1} + \sum_{k=1}^{p}\gamma_k(Y_{t-k} - \beta X_{t-k-1})\right\}\left\{\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} + \sum_{k=1}^{p}\gamma_k\frac{X_{t-k-1}}{\sqrt{1+X_{t-k-1}^2}}\right\},$$

and define the empirical likelihood function for (β,γ) as

$$L(\beta,\gamma) = \sup\left\{\prod_{t=1}^{n}(np_t) : p_1,\cdots,p_n \geq 0,\ \sum_{t=1}^{n}p_t = 1,\ \sum_{t=1}^{n}p_tZ_t(\beta,\gamma) = 0\right\}.$$

Since we are only interested in β, we consider the profile empirical likelihood function:

$$L^P(\beta) = \max_{\gamma\in\mathbb{R}^p} L(\beta,\gamma).$$

Theorem 1. *Suppose model (4.1) hold with either $|\phi| < 1$ independent of n or $\phi = 1 - \delta_0/n$ for some constant $\delta_0 \in R$. Further we assume $E|\varepsilon_t|^d < \infty$ for some $d > 2$, and the distribution of V_t is in the domain of attraction of a stable law with index $\alpha^* \in (0,2]$. Then $-2\log L^P(\beta_0) \xrightarrow{d} \chi^2(1)$ as $n \to \infty$, where β_0 denotes the true value of β.*

Next, we consider a predictive regression model with a linear time trend:

$$Y_t = \alpha_1 + \alpha_2 t + \beta X_{t-1} + U_t, \quad X_t = \theta + \phi X_{t-1} + e_t, \quad B(L)e_t = V_t, \quad \varepsilon_t = U_t + \sum_{j=1}^{p}\gamma_j U_{t-j}, \tag{4.5}$$

where $(\varepsilon_1,V_1),...,(\varepsilon_n,V_n)$ are independent and identically distributed random vectors. As explained in Zhu, Cai and Peng (2014 [1]), a directly application of the same or a similar weighted idea fails due to a degenerate limit. Instead, Zhu, Cai and Peng (2014 [1]) proposed to first split data into two parts and then to apply the empirical likelihood method to some weighted score equations based on the differences constructed from these two sub-samples.

Here one may need to split the sample into three parts due to the linear trend. Instead of splitting the data, we employ the trick of adding pseudo samples proposed by Li, Chan and Peng (2014 [93]) and Hill, Li and Peng (2014 [91]) to achieve a uniform inference.

More specifically, put

$$\bar{Z}_t(\beta,\alpha_1,\alpha_2,\gamma) := \left[\bar{Z}_{t,1}(\beta,\alpha_1,\alpha_2,\gamma),\cdots,\bar{Z}_{t,p+3}(\beta,\alpha_1,\alpha_2,\gamma)\right]^T \in \mathbb{R}^{p+3},$$

where

$$\begin{aligned}\bar{Z}_{t,j}(\beta,\alpha_1,\alpha_2,\gamma) &= \left\{Y_t-\alpha_1-\alpha_2 t-\beta X_{t-1}+\sum_{k=1}^{p}\gamma_k(Y_{t-k}-\alpha_1-\alpha_2(t-k)-\beta X_{t-k-1})\right\}\\ &\quad\times(Y_{t-j}-\alpha_1-\alpha_2(t-j)-\beta X_{t-j-1})\end{aligned}$$

for $j=1,\cdots,p$, and

$$\bar{Z}_{t,p+1}(\beta,\alpha_1,\alpha_2,\gamma) = Y_t-\alpha_1-\alpha_2 t-\beta X_{t-1}+\sum_{k=1}^{p}\gamma_k\left(Y_{t-k}-\alpha_1-\alpha_2(t-k)-\beta X_{t-k-1}\right),$$

$$\bar{Z}_{t,p+2}(\beta,\alpha_1,\alpha_2,\gamma) = \left\{Y_t-\alpha_1-\alpha_2 t-\beta X_{t-1}+\sum_{k=1}^{p}\gamma_k(Y_{t-k}-\alpha_1-\alpha_2(t-k)-\beta X_{t-k-1})\right\}t$$

$$\begin{aligned}\bar{Z}_{t,p+3}(\beta,\alpha_1,\alpha_2,\gamma) &= \left\{Y_t-\alpha_1-\alpha_2 t-\beta X_{t-1}+\sum_{k=1}^{p}\gamma_k(Y_{t-k}-\alpha_1-\alpha_2(t-k)-\beta X_{t-k-1})\right\}\\ &\quad\times\left\{\frac{X_{t-1}}{(1+X_{t-1}^2)^{\delta}}+\sum_{j=1}^{p}\gamma_j\frac{X_{t-j-1}}{(1+X_{t-j-1}^2)^{\delta}}\right\}+W_t \text{ for some } \delta>0.\end{aligned}$$

Note that we drop out the factor $1+\sum_{k=1}^{p}\gamma_k$ in $\bar{Z}_{t,p+1}$ and replace the factor $t+\sum_{k=1}^{p}\gamma_k(t-k)$ by t in $\bar{Z}_{t,p+2}$ since solving the score equations is invariant to these changes. The $W_t's$ are simulated independent and identically distributed random variables with $N(0,\bar{\sigma}^2)$, and $\bar{\sigma}^2 \geq 0$ is chosen not to be larger than $E(e_t^2)$. In order to avoid the effect of a random seed in generating $W_t's$, we use $W_t = 1/\sqrt{1000}\sum_{i=1}^{1000}W_{t,i}$ in our simulation study, where the $W_{t,i}'s$ for $t=1,\cdots,n$, and $i=1,\cdots,1000$ are a random sample from $N(0,\bar{\sigma}^2)$.

Note that when $\bar{\sigma}=0$, $\delta=1/2$ and $|X_t| \xrightarrow{p} \infty$, the joint limit of $1/\sqrt{n}\sum_{t=1}^{n}\bar{Z}_{t,p+3}(\beta,\alpha_1,\alpha_2,\gamma)$

and $1/\sqrt{n}\sum_{t=1}^{n}\bar{Z}_{t,p+1}(\beta,\alpha_1,\alpha_2,\gamma)$ is no longer normally distributed, which makes the application of the empirical likelihood method fail. This is why we need to add the pseudo sample W_t here to achieve uniform inference. Based on arguments presented in Li, Chan and Peng (2014 [93]) and Hill and Peng (2014 [91]), in the nonstationary case a choice of $\delta > 1/2$ makes $\sum_{t=1}^{n}\bar{Z}_{t,p+3}(\beta_0,\alpha_{0,1},\alpha_{0,2},\gamma_0)$ asymptotically equivalent to $\sum_{t=1}^{n}W_t$, while small $\delta \le 1$ allows $\sum_{t=1}^{n}\bar{Z}_{t,p+3}(\beta,\alpha_{0,1},\alpha_{0,2},\gamma_0)$ to better detect departures from β_0. Here $\beta_0,\alpha_{0,1},\alpha_{0,2},\gamma_0$ denote the true values of $\beta,\alpha_1,\alpha_2,\gamma$, respectively. Given the above arguments, we therefore enforce $\delta \in (1/2, 1]$ to balance power and size, and in practice simply use $\delta = 0.75$.

Finally we define the empirical likelihood function for $(\beta,\alpha_1,\alpha_2,\gamma)$ based on $\{\bar{Z}_t(\beta,\alpha_1,\alpha_2,\gamma)\}_{t=1}^{n}$ as

$$\bar{L}(\beta,\alpha_1,\alpha_2,\gamma)=\sup\left\{\prod_{t=1}^{n}(np_t): p_1,\cdots,p_n\ge 0,\ \sum_{t=1}^{n}p_t=1,\ \sum_{t=1}^{n}p_t\bar{Z}_t\left(\beta,\alpha_1,\alpha_2,\gamma\right)=0\right\},$$

and as before, for obtaining an interval for β we consider the profile empirical likelihood function

$$\bar{L}^P(\beta)=\max_{(\alpha_1,\alpha_2,\gamma^T)^T\in\mathbb{R}^{p+2}}\bar{L}(\beta,\alpha_1,\alpha_2,\gamma).$$

Theorem 2. *Suppose model (4.5) hold with either $|\phi|<1$ independent of n or $\phi = 1-\delta_0/n$ for some constant $\delta_0\in R$. Further we assume $E|\varepsilon_t|^d<\infty$ for some $d>2$, and the distribution of V_t is in the domain of attraction of a stable law with index $\alpha^*\in(0,2]$. Then $-2\log\bar{L}^P(\beta_0)\xrightarrow{d}\chi^2(1)$ as $n\to\infty$.*

Remark 1. When we consider model (4.5) with a constant trend, i.e., $\alpha_2 = 0$ in (4.5) is known, the above Theorem 2 still holds if the term $\bar{Z}_{t,p+2}$ is removed and α_2 is replaced by zero.

4.2.2 Jackknife Empirical Likelihood

The above profile empirical likelihood methods become computationally intensive when p is large. In order to reduce computational time, one may estimate γ first by solving (4.2), which results in an explicit function of β, and then apply the empirical likelihood method to (4.4) with γ replaced by the obtained estimator. However, this does not lead to a chi-squared limit due to the plug-in estimator. Recently a jackknife empirical likelihood method was proposed by Jing, Yuan and Zhou (2009 [94]) to deal with non-linear functionals, and Li, Peng and Qi (2011 [95]) employed this idea to reduce the computation of the empirical likelihood method based on estimating equations. Here, we employ the jackknife empirical likelihood method to reduce the computation of the above profile empirical likelihood method so as to give a unified interval estimation regardless of the process $\{X_t\}$ being stationary or non-stationary, or having a finite variance or an infinite variance.

We again first consider a simpler model, i.e., model (4.1). Let $\hat{\gamma}(\beta) = (\hat{\gamma}_1(\beta), \cdots, \hat{\gamma}_p(\beta))^T$ be, for arbitrary β, the solution to (4.2), and for each $i = 1, \cdots, n$ let $\hat{\gamma}^{(i)}(\beta) = (\hat{\gamma}_1^{(i)}(\beta), \cdots, \hat{\gamma}_p^{(i)}(\beta))^T$ be the solution to

$$\sum_{t=1, t\neq i}^{n} \left\{ Y_t - \beta X_{t-1} + \sum_{k=1}^{p} \gamma_k (Y_{t-k} - \beta X_{t-k-1}) \right\} (Y_{t-j} - \beta X_{t-j-1}) = 0 \quad \text{for} \quad j = 1, \cdots, p. \tag{4.6}$$

Next we define the pseudo sample as

$$\begin{aligned} Z_j^*(\beta) &= \sum_{t=1}^{n} \left\{ Y_t - \beta X_{t-1} + \sum_{k=1}^{p} \hat{\gamma}_k(\beta)(Y_{t-k} - \beta X_{t-k-1}) \right\} \\ &\qquad \times \left\{ \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} + \sum_{k=1}^{p} \hat{\gamma}_k(\beta) \frac{X_{t-k-1}}{\sqrt{1+X_{t-k-1}^2}} \right\} \\ &\quad - \sum_{t=1, t\neq j}^{n} \left\{ Y_t - \beta X_{t-1} + \sum_{k=1}^{p} \hat{\gamma}_k^{(j)}(\beta)(Y_{t-k} - \beta X_{t-k-1}) \right\} \\ &\qquad \times \left\{ \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} + \sum_{k=1}^{p} \hat{\gamma}_k^{(j)}(\beta) \frac{X_{t-k-1}}{\sqrt{1+X_{t-k-1}^2}} \right\} \end{aligned}$$

for $j = 1, \cdots, n$. Based on this pseudo sample, the jackknife empirical likelihood function for β is defined as

$$L^*(\beta) = \sup \left\{ \prod_{i=1}^{n} (np_i) : p_1 \geq 0, \cdots, p_n \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i Z_i^*(\beta) = 0 \right\}.$$

Theorem 3. *Suppose model (4.1) hold with either* $|\phi| < 1$ *independent of* n *or* $\phi = 1 - \delta_0/n$ *for some constant* $\delta_0 \in R$. *Further we assume* $E|\varepsilon_t|^d < \infty$ *for some* $d > 4$, *and the distribution of* V_t *is in the domain of attraction of a stable law with index* $\alpha^* \in (0, 2]$. *Then* $-2 \log L^*(\beta_0) \xrightarrow{d} \chi^2(1)$ *as* $n \to \infty$.

Next, we consider a model with a linear time trend, i.e., model (4.5). Define the parameter subset

$$\theta = [\alpha_1, \alpha_2, \beta]^T .$$

Let $\bar{\gamma}(\theta) = (\bar{\gamma}_1(\theta), \cdots, \bar{\gamma}_p(\theta))^T$ denote, for arbitrary θ, the solution to

$$\sum_{t=1}^{n} \bar{Z}_{t,j}(\theta, \gamma) = 0 \quad \text{for} \quad j = 1, \cdots, p, \tag{4.7}$$

and let $\bar{\gamma}^{(i)}(\theta) = (\bar{\gamma}_1^{(i)}(\theta), \cdots, \bar{\gamma}_p^{(i)}(\theta))^T$ for $i = 1, \cdots, n$ denote the solution to

$$\sum_{t=1, t \neq i}^{n} \bar{Z}_{t,j}(\theta, \gamma) = 0 \quad \text{for} \quad j = 1, \cdots, p. \tag{4.8}$$

Then a jackknife pseudo sample is obtained as

$$\bar{Z}_{i,k}^*(\theta) = \sum_{t=1}^{n} \bar{Z}_{t,p+k}(\theta, \bar{\gamma}) - \sum_{t=1, t \neq i}^{n} \bar{Z}_{t,p+k}(\theta, \bar{\gamma}^{(i)}),$$

for $i = 1, \cdots, n$ and $k = 1, 2, 3$. Put

$$\bar{Z}_i^*(\theta) = (\bar{Z}_{i,1}^*(\theta), \bar{Z}_{i,2}^*(\theta), \bar{Z}_{i,3}^*(\theta)),$$

and define the jackknife empirical likelihood function for $(\beta, \alpha_1, \alpha_2)$ as

$$\bar{L}^*(\beta, \alpha_1, \alpha_2) = \bar{L}^*(\theta) = \sup \left\{ \prod_{i=1}^{n} (np_i) : p_1 \geq 0, \cdots, p_n \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \bar{Z}_i^*(\theta) = 0 \right\}.$$

Since we are only interested in β, we consider the following profile jackknife empirical likelihood function

$$\bar{L}^{*P}(\beta) = \max_{(\alpha_1,\alpha_2)^T \in \mathbb{R}^2} \bar{L}^*(\beta, \alpha_1, \alpha_2).$$

Theorem 4. *Suppose model (4.5) hold with either $|\phi| < 1$ independent of n or $\phi = 1 - \delta_0/n$ for some constant $\delta_0 \in R$. Further we assume $E|\varepsilon_t|^d < \infty$ for some $d > 4$, and the distribution of V_t is in the domain of attraction of a stable law with index $\alpha^* \in (0, 2]$. Then $-2 \log \bar{L}^{*P}(\beta_0) \xrightarrow{d} \chi^2(1)$ as $n \to \infty$.*

Remark 2. *Based on the above theorems, confidence intervals for β with level a can be obtained as*

$$I_a = \{\beta : -2 \log L^P(\beta) \le \chi^2_{1,a}\}, \quad \bar{I}_a = \{\beta : -2 \log \bar{L}^P(\beta) \le \chi^2_{1,a}\},$$

$$I^*_a = \{\beta : -2 \log L^*(\beta) \le \chi^2_{1,a}\}, \quad \bar{I}^*_a = \{\beta : -2 \log \bar{L}^{*P}(\beta) \le \chi^2_{1,a}\},$$

where $\chi^2_{1,a}$ is the a-th quantile of a chi-squared distribution with one degree of freedom.

4.3 Simulation study

In this section we examine the finite sample behavior of the proposed methods for models (4.1), (4.5) and (4.5) with known $\alpha_2 = 0$.

Let $\{(\varepsilon^*_i, V^*_i)\}_{i=1}^n$ be a random sample from a bivariate normal distribution with zero means, one variances and 0.5 correlation coefficient. Let $\{T_i\}_{i=1}^n$ be a random sample from a t-distribution with degrees freedom ν and independent of $\{(\varepsilon^*_i, V^*_i)\}_{i=1}^n$. Then we take $\varepsilon_i = \varepsilon^*_i$ and $V_i = V^*_i + T_i$ for $i = 1, \cdots, n$ throughout, and draw $10,000$ random samples with size $n = 50$ and 200 from either model (4.1) or (4.5), with $\phi \in \{.9, .99, 1\}$, $p = 5$ with $\gamma = \left(\binom{p}{1}0.4, \binom{p}{2}0.4^2, \cdots, \binom{p}{p}0.4^p\right)$, $q = 1$, $b_1 = -0.4$, $\theta = 1$, $\beta = 0$ and $\nu = 1.5$ or 3. In the case of model (4.5) we use $\alpha_1 = .2$ and $\alpha_2 \in \{0, .2\}$. The added pseudo sample $\{W_t\}_{t=1}^n$ is computed using $W_t = 1/\sqrt{1000}\sum_{i=1}^{1000} W_{t,i}$, where $W'_{t,i}s$ are a random sample from $N(0, .5)$. We employ the R package 'emplik' to compute the empirical likelihood function and then use the R package 'nlm' to calculate the profile empirical likelihood function. For using 'nlm' to compute coverage probabilities for model (4.1), we choose the initial value for γ by minimizing

$$\sum_{t=1}^{n}\{Y_t-\beta X_{t-1}+\sum_{j=1}^{p}\gamma_j(Y_{t-j}-\beta X_{t-j-1})\}^2$$

for each fixed β. However, for using 'nlm' to compute coverage probabilities for model (4.5), we use the following procedure to choose initial values for α_1,α_2 and γ. We first minimize $\sum_{t=1}^{n}\{Y_t-\alpha_1-\alpha_2 t-\beta X_{t-1}\}^2$ for each fixed β to obtain $\tilde{\alpha}_i(\beta)$, and then minimize

$$\sum_{t=1}^{n}\left\{Y_t-\tilde{\alpha}_1(\beta)-\tilde{\alpha}_2(\beta)\,t-\beta X_{t-1}+\sum_{k=1}^{p}\gamma_k(Y_{t-k}-\tilde{\alpha}_1(\beta)-\tilde{\alpha}_2(\beta)\,(t-k)-\beta X_{t-k-1})\right\}^2$$

with respect to γ to achieve $\tilde{\gamma}(\beta)$. Finally the initial values for α_1 and α_2 are chosen to minimize

$$\sum_{t=1}^{n}\{Y_t-\alpha_1-\alpha_2 t-\beta X_{t-1}+\sum_{k=1}^{p}\tilde{\gamma}_k(\beta)(Y_{t-k}-\alpha_1-\alpha_2(t-k)-\beta X_{t-k-1})\}^2$$

for each fixed β, say $\hat{\alpha}_1(\beta)$ and $\hat{\alpha}_2(\beta)$, and the initial value of γ is chosen to minimize

$$\sum_{t=1}^{n}\{Y_t-\hat{\alpha}_1(\beta)-\hat{\alpha}_2(\beta)t-\beta X_{t-1}+\sum_{k=1}^{p}\gamma_k(Y_{t-k}-\hat{\alpha}_1(\beta)-\hat{\alpha}_2(\beta)(t-k)-\beta X_{t-k-1})\}^2$$

for each fixed β. Note that we only need to compute the profile (jackknife) empirical likelihood functions at $\beta=\beta_0$ for calculating the coverage probabilities.

Coverage probabilities are reported in Tables 1–3. We observe from these tables that i) coverage probabilities for $n=200$ are much closer to the nominal level than those for $n=50$; ii) the profile empirical likelihood method is worse than the jackknife empirical likelihood method for most cases; and iii) the jackknife empirical likelihood method gives more accurate coverage probabilities for model (4.5) with $\alpha_2=0$ known than that with unknown α_2. In conclusion, the proposed methods perform quite well for all considered cases especially for $n=200$.

4.4 Data Analysis

In this section, we re-visit the data sets analyzed by Campbell ad Yogo (2006 [36]) and Zhu, Cai and Peng (2014 [1]), where the monthly CRSP value-weighted index (1926:12–2002:12) is used as predictable variable Y_t, and either the log dividend-price ratio (ldp) or the log earnings-price ratio (lep) is treated as predicting variable X_t. We fit this data set to model (4.5) with $\alpha_2 = 0$ known.

First we estimate α_1, β, γ by minimizing the following least squares

$$\sum_{t=1}^{n}\{Y_t - \alpha_1 - \beta X_{t-1} + \sum_{k=1}^{p}\gamma_k(Y_{t-k} - \alpha_1 - \beta X_{t-k-1})\}^2,$$

say $\tilde{\alpha}_1^{(p)}, \tilde{\beta}^{(p)}, \tilde{\gamma}^{(p)}$. Based on these estimators, we estimate ε_t by

$$\tilde{\varepsilon}_t^{(p)} = Y_t - \tilde{\alpha}_1^{(p)} - \tilde{\beta}^{(p)}X_{t-1} + \sum_{k=1}^{p}\tilde{\gamma}_k^{(p)}(Y_{t-k} - \tilde{\alpha}_1^{(p)} - \tilde{\beta}^{(p)}X_{t-k-1})$$

for $t = 1, \cdots, n$. Hence, the standard deviation of ε_t is estimated by

$$\tilde{\sigma}^{(p)} = \{\frac{1}{n}\sum_{t=1}^{n}(\tilde{\varepsilon}_t^{(p)} - \frac{1}{n}\sum_{i=1}^{n}\tilde{\varepsilon}_i^{(p)})^2\}^{1/2}.$$

Note that $\tilde{\varepsilon}_t^{(p)}$ and $\tilde{\sigma}^{(p)}$ become an estimator for U_t and the standard deviation of U_t, respectively when $p = 0$. In Figures 1 and 2, we plot the autocorrelation functions by applying the 'acf' function in R to $\{\tilde{\varepsilon}_t^{(p)}\}_{t=1}^n$ with $p = 0, 5, 10, 20$, where the plots with $p = 0$ clearly show that independence assumption for $U_t's$ is questionable. When p becomes larger, the dependence among $\varepsilon_t's$ tends to be weaker.

Next we employ the proposed jackknife empirical likelihood method to test $H_0 : \beta = 0$ against $H_a : \beta \neq 0$. To reduce the effect of the added pseudo sample in the proposed jackknife empirical likelihood methods, we repeat the test $1,000$ times and report the average of these obtained $1,000$ P-values in Table 4, which concludes that the null hypothesis of no predictability can not be rejected for these two predicting variables. However, both Campbell and Yogo (2006 [36]) and Zhu, Cai and Peng (2014 [1]) rejected the null hypothesis of no predictability when the predicting variable is the log earnings-price ratio, which may be due to the ignored dependence among errors.

When the predicting variable is nearly integrated, the proposed methods mainly depend on the behavior of $\varepsilon_t's$. From Figures 1 and 2, the estimated $\varepsilon_t's$ for both predicting variables have a similar pattern of autocorrelation function. Hence, the P-values of the proposed jackknife empirical likelihood method for both predicting variables may be similar, which is supported by Table 4.

Table (4.1) *No trend.* Coverage probabilities based on Theorem 1 and Theorem 3 are reported for model (4.1) with levels 0.9 and 0.95, where intervals are defined in Remark 2.

(ϕ, n)	$I_{0.9}$ $\nu = 1.5$	$I^*_{0.9}$ $\nu = 1.5$	$I_{0.95}$ $\nu = 1.5$	$I^*_{0.95}$ $\nu = 1.5$	$I_{0.9}$ $\nu = 3$	$I^*_{0.9}$ $\nu = 3$	$I_{0.95}$ $\nu = 3$	$I^*_{0.95}$ $\nu = 3$
$(0.9, 50)$	0.8799	0.8874	0.9337	0.9430	0.8828	0.8914	0.9353	0.9454
$(0.99, 50)$	0.8740	0.8853	0.9274	0.9393	0.8796	0.8929	0.9368	0.9471
$(1, 50)$	0.8729	0.8836	0.9277	0.9395	0.8784	0.8914	0.9347	0.9451
$(0.9, 200)$	0.9055	0.9085	0.9534	0.9581	0.8932	0.8978	0.9445	0.9478
$(0.99, 200)$	0.8917	0.8959	0.9470	0.9521	0.8912	0.8959	0.9470	09512
$(1, 200)$	0.8920	0.8965	0.9428	0.9474	0.8887	0.8949	0.9453	0.9511

Table (4.2) *Linear time trend.* Coverage probabilities based on Theorem 2 and Theorem 4 are reported for model (4.5) with levels 0.9 and 0.95, where intervals are defined in Remark 2.

(ϕ, n)	$\bar{I}_{0.9}$ $\nu = 1.5$	$\bar{I}^*_{0.9}$ $\nu = 1.5$	$\bar{I}_{0.95}$ $\nu = 1.5$	$\bar{I}^*_{0.95}$ $\nu = 1.5$	$\bar{I}_{0.9}$ $\nu = 3$	$\bar{I}^*_{0.9}$ $\nu = 3$	$\bar{I}_{0.95}$ $\nu = 3$	$\bar{I}^*_{0.95}$ $\nu = 3$
$(0.9, 50)$	0.8501	0.8578	0.9110	0.9199	0.8412	0.8483	0.9062	0.9132
$(0.99, 50)$	0.8546	0.8645	0.9210	0.9261	0.8482	0.8563	0.9098	0.9190
$(1, 50)$	0.8598	0.8698	0.9218	0.9288	0.8541	0.8656	0.9162	0.9249
$(0.9, 200)$	0.8810	0.8834	0.9405	0.9443	0.8830	0.8861	0.9382	0.9400
$(0.99, 200)$	0.8846	0.8887	0.9385	0.9408	0.8803	0.8836	0.9337	0.9370
$(1, 200)$	0.8929	0.8950	0.9448	0.9468	0.8820	0.8841	0.9356	0.9382

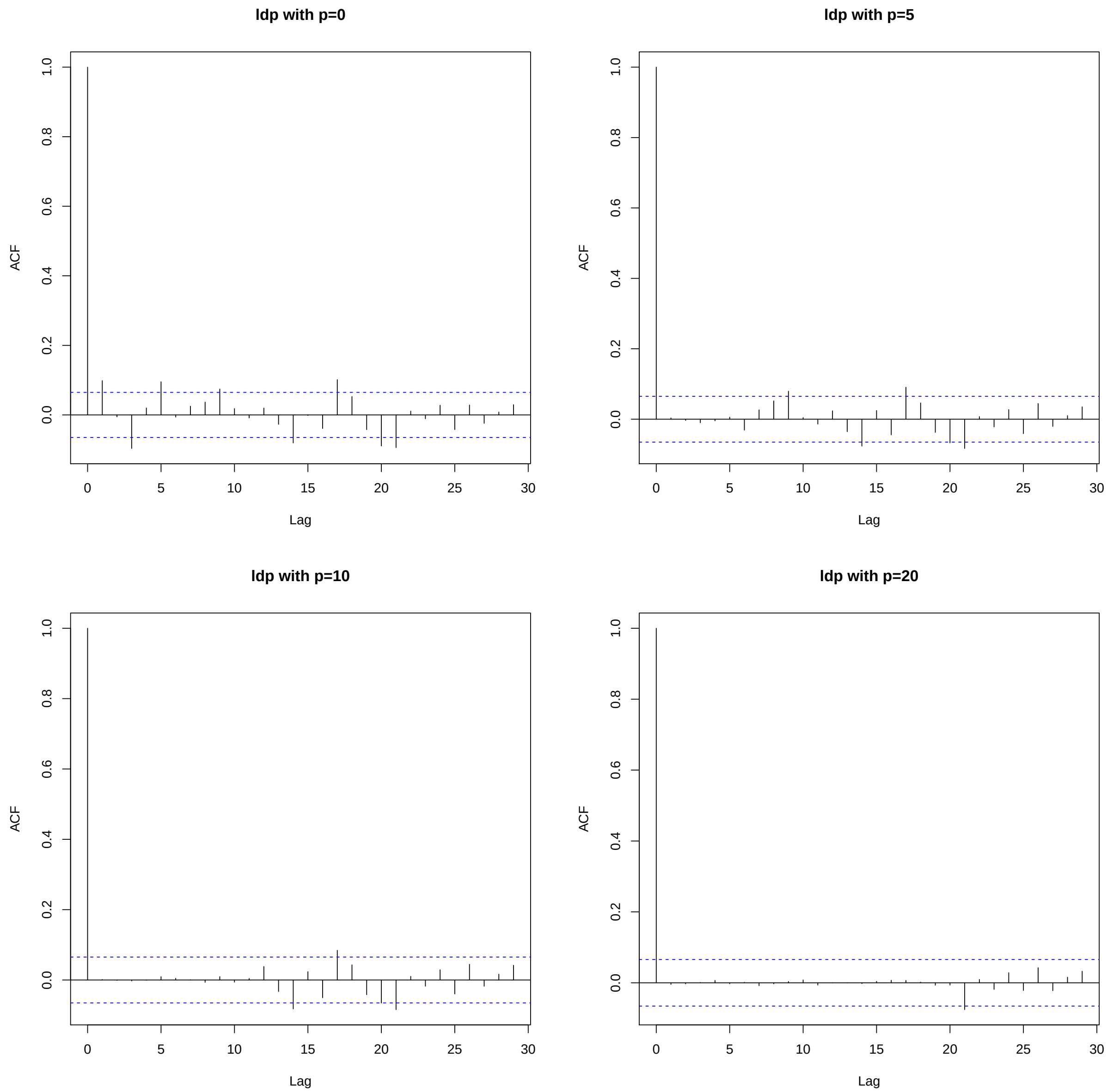


Figure (4.1) Autocorrelation function is plotted based on estimated $\varepsilon_t's$ when the predicting variable is the log dividend-price ratio.

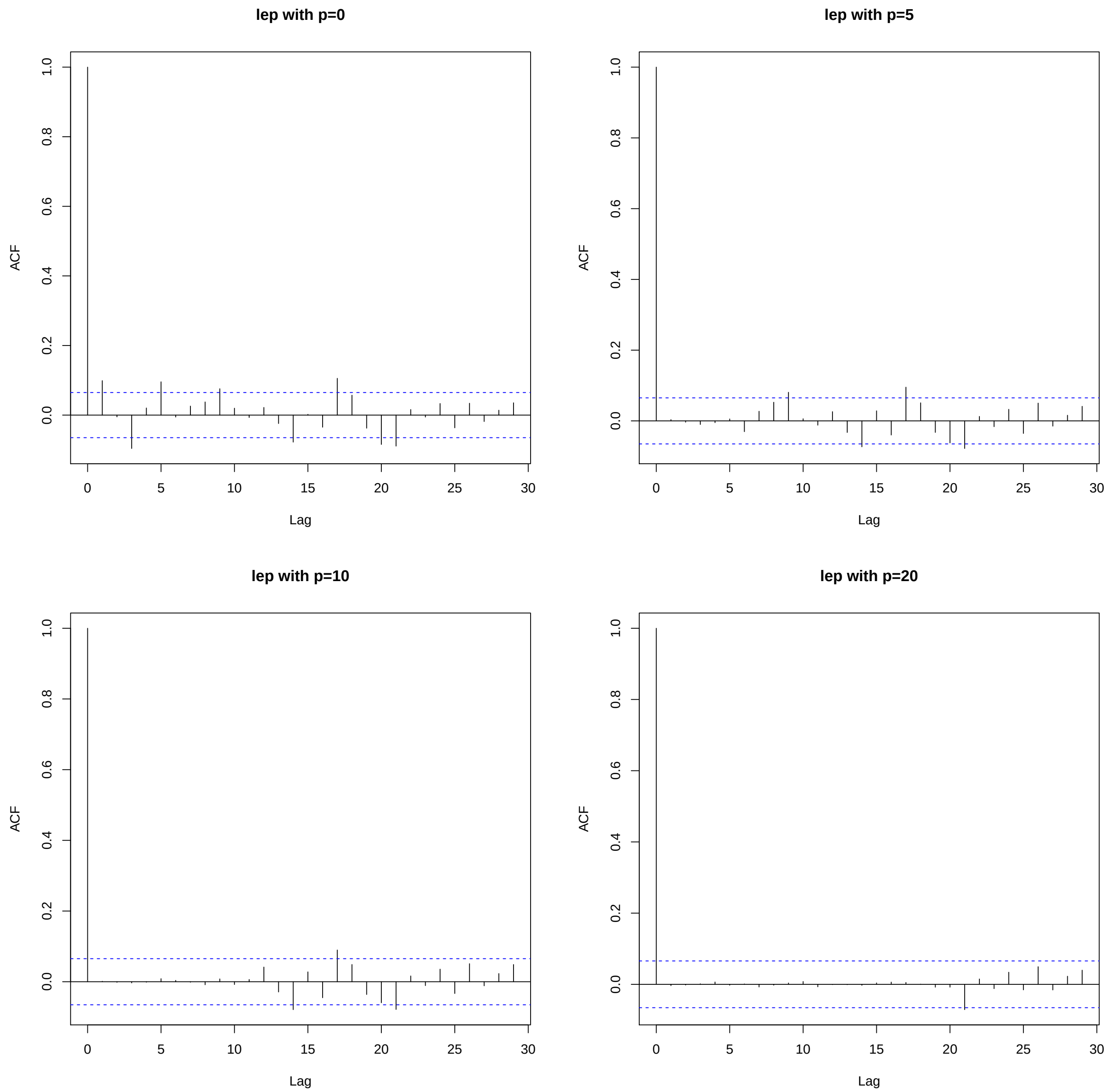


Figure (4.2) Autocorrelation function is plotted based on estimated $\varepsilon_t' s$ when the predicting variable is the log earnings-price ratio.

Table (4.3) *Constant trend.* Coverage probabilities based on Theorem 2 and Theorem 4 are reported for model (4.5) with $\alpha_2 = 0$ known and levels 0.9 and 0.95, where intervals are defined in Remark 2.

(ϕ, n)	$\bar{I}_{0.9}$ $\nu = 1.5$	$\bar{I}^*_{0.9}$ $\nu = 1.5$	$\bar{I}_{0.95}$ $\nu = 1.5$	$\bar{I}^*_{0.95}$ $\nu = 1.5$	$\bar{I}_{0.9}$ $\nu = 3$	$\bar{I}^*_{0.9}$ $\nu = 3$	$\bar{I}_{0.95}$ $\nu = 3$	$\bar{I}^*_{0.95}$ $\nu = 3$
$(0.9, 50)$	0.8593	0.8658	0.9191	0.9274	0.8549	0.8672	0.9175	0.9262
$(0.99, 50)$	0.8684	0.8752	0.9246	0.9334	0.8612	0.8715	0.9169	0.9281
$(1, 50)$	0.8678	0.8738	0.9248	0.9337	0.8632	0.8716	0.9181	0.9300
$(0.9, 200)$	0.8920	0.8957	0.9446	0.9467	0.8925	0.8964	0.9447	0.9475
$(0.99, 200)$	0.8889	0.8916	0.9406	0.9429	0.8860	0.8906	0.9427	0.9460
$(1, 200)$	0.8874	0.8896	0.9428	0.9456	0.8859	0.8898	0.9399	0.9432

Table (4.4) *P-values.* The average of P-Values is reported by repeating $1,000$ times of the proposed jackknife empirical likelihood method for testing $H_0 : \beta = 0$ against $H_a : \beta \neq 0$ based on model (4.5) with $\alpha_2 = 0$ known.

	ldp $\bar{\sigma} = \tilde{\sigma}^{(p)}$	ldp $\bar{\sigma} = \tilde{\sigma}^{(p)}/2$	lep $\bar{\sigma} = \tilde{\sigma}^{(p)}$	lep $\bar{\sigma} = \tilde{\sigma}^{(p)}/2$
$p = 0$	0.4901	0.4894	0.4894	0.4878
$p = 5$	0.4935	0.4936	0.4930	0.4923
$p = 10$	0.4941	0.4948	0.4938	0.4945
$p = 20$	0.4932	0.4949	0.4931	0.4951

Chapter 5

INFERENCE FOR TAIL INDEX OF GARCH(1,1) MODEL AND AR(1) MODEL WITH ARCH(1) ERRORS

This chapter is based on the following revised paper by R. Zhang, C. Li and L. Peng (2015), *Inference for Tail Index of GARCH(1,1) Model and AR(1) Model with ARCH(1) Errors Econometric Reviews.*

5.1 Outline

As we stated in Chapter 1 1.3, in this chapter, for tail index of GARCH(1,1) Model and AR(1) Model with ARCH(1) Errors, we will illustrate our methods in details. This chapter is organized as follows. The proposed methodologies and their asymptotic results are presented in Section 5.2. Section 5.3 presents a simulation study and a data analysis. Some conclusions are given in Section 5.4. All technical proofs are put in Appendix A.

5.2 Models, Methodologies and Theoretical Results

5.2.1 Heavy Tailed GARCH(1, 1) Model

Being a benchmark of GARCH family, GARCH(1, 1) model is simply used to capture the heteroscedastic and heavy-tailed phenomena in financial returns, which is defined as

$$Y_t = \sigma_t^* \varepsilon_t^*, \ (\sigma_t^*)^2 = \omega^* + a^*(\sigma_{t-1}^*)^2 + b^* Y_{t-1}^2, \tag{5.1}$$

where $\omega^* > 0$, $a^* \geq 0$, $b^* \geq 0$ and $\{\varepsilon_t^*\}$ is a sequence of independent and identically distributed random variables with zero mean and unit variance. For some general studies and applications of GARCH models in financial econometrics, we refer to Tayor (2005 [37]) and Francq and Zakoïan (2010 [96]). For model (5.1), it is known that, under some conditions,

Y_t has a heavy tail with index $\alpha > 0$. More specifically, it follows from Basrak, Davis and Mikosch (2002 [41]) that

$$P(|Y_t| > x) = cx^{-\alpha}\{1 + o(1)\} \text{ for some } c > 0 \text{ as } x \to \infty, \tag{5.2}$$

and the tail index α is determined by

$$\mathrm{E}\{a^* + b^*(\varepsilon_t^*)^2\}^{\alpha/2} = 1. \tag{5.3}$$

Note that equations (5.2) and (5.3) can be derived from Kesten (1973 [55]) and Goldie (1991 [97]) too. When $\mathrm{E}|\varepsilon_t^*|^\delta < \infty$ for some $\delta > \max\{4, 2\alpha\}$, one can estimate the nuisance parameters $\theta^* = (\omega^*, a^*, b^*)^T$ by the QMLE (say $\hat{\theta}^* = (\hat{\omega}^*, \hat{a}^*, \hat{b}^*)^T$) and then estimate the tail index α by solving the following estimating equation:

$$\frac{1}{n}\sum_{t=1}^{n}\{\hat{a}^* + \hat{b}^*(\hat{\varepsilon}_t^*)^2\}^{\alpha/2} = 1, \tag{5.4}$$

where $\hat{\varepsilon}_t^* = Y_t/\hat{\sigma}_t^*$ and $\hat{\sigma}_t^*$ is an estimator of σ_t^* with θ^* being replaced by $\hat{\theta}^*$, see Berkes, Horváth and Kokoszka (2003 [61]) for the asymptotic distribution of the above estimator and Chan, Peng and Zhang (2012 [62]) for a profile empirical likelihood inference based on the above estimation procedure.

Note that $\delta > 4$ ensures that the QMLE $\hat{\theta}^*$ has a normal limit, and $\delta > 2\alpha$ ensures the asymptotic normality for estimating α via solving $n^{-1}\sum_{t=1}^{n}\{a^* + b^*(\epsilon_t^*)^2\}^{\alpha/2} = 1$. Therefore the condition of $\delta > 2\alpha$ can not be relaxed. However, we may be able to allow $\mathrm{E}|\varepsilon_t^*|^4 = \infty$ by using some different estimate for parameters θ^* such as least absolute deviations estimate, which generally requires to reparameterize model (5.1). Issues on reparameterization for GARCH sequences are discussed in Fan, Qi and Xiu (2014 [98]).

Assume the unknown median of $(\varepsilon_t^*)^2$ is $d > 0$ and put $\varepsilon_t = \varepsilon_t^*/\sqrt{d}$. Then the median of $\log \varepsilon_t^2$ becomes $\log\{\text{median}((\varepsilon_t^*)^2/d)\} = 0$. Furthermore, model (5.1) and equation (5.3) can be written as

$$Y_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \omega + a\sigma_{t-1}^2 + bY_{t-1}^2 \tag{5.5}$$

and

$$E\{a + b\epsilon_t^2\}^{\alpha/2} = 1, \tag{5.6}$$

where $\sigma_t = \sqrt{d}\sigma_t^*$, $\omega = d\omega^*$, $a = a^*$ and $b = db^*$. It is clear that the estimating equation for the tail index α remains unchanged. Therefore we propose to first apply the least absolute deviations estimate in Peng and Yao (2003 [99]) to (5.5) and then to estimate the tail index α via (5.6) so as to relax the moment condition on ε_t^* or equivalently on ε_t.

More specifically, for any $\theta = (\omega, a, b)^T$, by the recursion of (5.5), the conditional variance $\sigma_t^2 = \sigma_t^2(\theta)$ can be represented as follows:

$$\sigma_t^2(\theta) \quad = \quad \omega + a\sigma_{t-1}^2(\theta) + bY_{t-1}^2 = \frac{\omega(1-a^t)}{1-a} + \sum_{k=0}^{t-1} ba^k Y_{t-1-k}^2 + a^t\sigma_0^2(\theta). \tag{5.7}$$

Thus, given the observations $\{Y_1, Y_2, \cdots, Y_n\}$ and the initial value Y_0, we can estimate θ by the following LADE:

$$\hat{\theta}_{inital} = \arg\min_{\theta} \sum_{t=1}^{n} |\log Y_t^2 - \log \sigma_t^2(\theta)|. \tag{5.8}$$

However, since $\sigma_0^2(\theta)$ depends on the unobserved sample path $Y_{-1}, Y_{-2}, \ldots$, one cannot use the above expression of $\sigma_t^2(\theta)$ in practice. Instead, we consider the LADE based on a truncated version of $\sigma_t(\theta)$, which is

$$\hat{\theta} = \arg\min_{\theta} \sum_{t=1}^{n} |\log Y_t^2 - \log \bar{\sigma}_t^2(\theta)|, \tag{5.9}$$

where $\bar{\sigma}_t^2(\theta) = \omega(1-a^t)/(1-a) + b\sum_{0\le k\le t-1} a^k Y_{t-k-1}^2$. Using this LADE $\hat{\theta}$, we estimate α by solving

$$\frac{1}{n}\sum_{t=1}^{n}(\hat{a} + \hat{b}\bar{\varepsilon}_t^2(\hat{\theta}))^{\alpha/2} = 1,$$

where $\bar{\varepsilon}_t^2(\hat{\theta}) = Y_t^2/\bar{\sigma}_t^2(\hat{\theta})$. Denote this estimator by $\hat{\alpha}$. For deriving the asymptotic limit of $\hat{\alpha}$, we need some regularity conditions:

Condition 1. $\mathrm{E}\log(a_0^* + b_0^*(\epsilon_t^*)^2) < 0$ (i.e., $\mathrm{E}\log(a_0 + b_0\epsilon_t^2) < 0$) and $\mathrm{E}|\varepsilon_t^*|^{\delta_0} < \infty$ (i.e., $\mathrm{E}|\varepsilon_t|^{\delta_0} < \infty$) for some $\delta_0 > \max\{2, 2\alpha_0\}$, where $\theta_0 = (\omega_0, a_0, b_0)^T$, $\theta_0^* = (\omega_0^*, a_0^*, b_0^*)^T$ and α_0 denote the true values of θ, θ^* and α respectively.

Condition 2. $(\varepsilon_t^*)^2$ has an unknown median $d > 0$ and a continuous density at d, i.e., $\log\{\varepsilon_t^2\}$ has median zero and its density $f(x)$ is continuous at zero.

Remark 3. $\mathrm{E}\log(a_0 + b_0\epsilon_t^2) < 0$ *in Condition 1 is a sufficient and necessary condition for the existence of a stationary solution of* σ_t^2 *(see Nelson (1990)). Further, Condition 1 and (5.3) imply that* b_0 *can not be zero, as a result, we have* $a_0 < 1$*. Condition 2 is a standard condition for a LADE, which is the same as that in Peng and Yao (2003 [99]).*

Remark 4. *When* $a_0 + b_0 < 1$*, it is known that* Y_t *has a finite variance (see Fan and Yao (2003 [100])), i.e.,* $\alpha_0 > 2$*. Therefore Condition 1 implies* ϵ_t^* *has a finite fourth moment in case of* $a_0 + b_0 < 1$*. In order to consider the case of infinite fourth moment for errors, one has to study the case of* $a + b \geq 1$*.*

Theorem 5. *Assume Conditions 1 and 2 hold for model (5.1). Then, as* $n \to \infty$

$$\sqrt{n}(\hat{\alpha} - \alpha_0) \xrightarrow{d} N(0, \gamma_{\alpha_0}^2), \tag{5.10}$$

where

$$\begin{aligned}\gamma_{\alpha_0}^2 &= \{4A_0^2 f^2(0)\}^{-1}(\mu_1, \mu_2, \mu_3)\Omega^{-1}(\mu_1, \mu_2, \mu_3)^T + 4\{A_0^2\}^{-1}\mathrm{E}[(a_0 + b_0\varepsilon_1^2)^{\frac{\alpha_0}{2}} - 1]^2 \\ &\quad + 2\{A_0^2 f(0)\}^{-1}(\mu_1, \mu_2, \mu_3)\Omega^{-1}\mathrm{E}\{A(1)[(a_0 + b_0\varepsilon_1^2)^{\frac{\alpha_0}{2}} - 1]\}\end{aligned}$$

with

$$\begin{cases} A_0 = \mathrm{E}[(a_0 + b_0\varepsilon_1^2)^{\frac{\alpha_0}{2}} \log(a_0 + b_0\varepsilon_1^2)], \\ e_0 = \alpha_0 \mathrm{E}[(a_0 + b_0\varepsilon_1^2)^{\alpha_0/2-1}\varepsilon_1^2], \\ \mu_1 = -\frac{b_0 e_0}{2}\mathrm{E}\frac{\partial \log \sigma_1^2(\theta_0)}{\partial w}, \\ \mu_2 = \alpha_0 \mathrm{E}[(a_0 + b_0\varepsilon_1^2)^{\frac{\alpha_0}{2}-1}] - \frac{b_0 e_0}{2}\mathrm{E}\frac{\partial \log \sigma_1^2(\theta_0)}{\partial a}, \\ \mu_3 = e_0 - \frac{b_0 e_0}{2}\mathrm{E}\frac{\partial \log \sigma_1^2(\theta_0)}{\partial b}, \\ \Omega = \mathrm{E}[A(1)A^T(1)], \\ A(t) = \left(\frac{\partial(\log \sigma_t^2(\theta))}{\partial \omega}, \frac{\partial(\log \sigma_t^2(\theta))}{\partial a}, \frac{\partial(\log \sigma_t^2(\theta))}{\partial b}\right)^T sgn\{\log(\varepsilon_t^2)\}, \end{cases}$$

and sgn denotes the sign function.

Remark 5. *Although the moment condition on errors depends on the unknown parameter* α_0*, this can be checked when the error distribution has heavy tails. More specifically one can simply compute the Hill's estimator based on estimated errors via quasi maximum likelihood estimators for parameters in the GARCH(1,1) model and then compare it with* $\max(2, 2\hat{\alpha})$*; see the data analysis in Section 3.*

To construct a confidence interval for the tail index α, an obvious approach is to estimate the asymptotic variance $\gamma_{\alpha_0}^2$. Due to the complexity of this asymptotic variance, one can simply employ a naive bootstrap method. However bootstrapping nonpivotal statistics is inefficient in general. Bootstrap method for a time series model is computationally intensive since one has to resample from the estimated errors and refit the time series model. Alternatively, we seek an empirical likelihood method to bypass estimating the asymptotic variance. A direct application of the empirical likelihood method to equation (5.6) with θ replaced by $\hat{\theta}$ cannot capture the variance of the plug-in estimator $\hat{\theta}$ since the asymptotic variance $\gamma_{\alpha_0}^2$ of the tail index estimator $\hat{\alpha}$ really depends on the asymptotic variances of $\hat{\theta}$. Hence, Wilks theorem fails for such a direct application of an empirical likelihood method. Instead we propose the following profile empirical likelihood method.

Note that the proposed LADE is a solution to the score equations

$$\sum_{t=1}^{n} \bar{Z}_{t,j}(\theta) = 0 \quad \text{for} \quad j = 2, 3, 4,$$

where

$$\begin{cases} \bar{Z}_{t,2}(\theta) = (\partial(\log \bar{\sigma}_t^2(\theta))/\partial\omega)\,\mathrm{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\}, \\ \bar{Z}_{t,3}(\theta) = (\partial(\log \bar{\sigma}_t^2(\theta))/\partial a)\,\mathrm{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\}, \\ \bar{Z}_{t,4}(\theta) = (\partial(\log \bar{\sigma}_t^2(\theta))/\partial b)\,\mathrm{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\}. \end{cases}$$

It follows from (5.6) that θ and α can be estimated simultaneously by solving the following equations

$$\sum_{t=1}^{n} \bar{Z}_{t,j} = 0 \quad \text{for} \quad j = 1, 2, 3, 4,$$

where $\bar{Z}_{t,1} := \bar{Z}_{t,1}(\theta, \alpha) = \{a + bY_t^2/\bar{\sigma}_t^2(\theta)\}^{\alpha/2} - 1$. This simultaneous estimation procedure motivates us to apply the empirical likelihood method to the above four equations and then profile the nuisance parameters θ. This is the so-called profile empirical likelihood method based on estimating equations proposed by Qin and Lawless (1994 [92]).

Put $\bar{Z}_t(\theta, \alpha) = (\bar{Z}_{t,1}^T(\theta, \alpha), \bar{Z}_{t,2}(\theta), \bar{Z}_{t,3}(\theta), \bar{Z}_{t,4}(\theta))^T$ for $t = 1, \ldots, n$, and define the empirical likelihood function of (θ, α) as

$$L(\theta, \alpha) = \sup\{\prod_{t=1}^{n}(np_t) : p_1 \geq 0, \ldots, p_n \geq 0, \sum_{t=1}^{n} p_t = 1, \sum_{t=1}^{n} p_t \bar{Z}_t(\theta, \alpha) = 0\}.$$

By virtue of the Lagrange multipliers, it is clear that $p_t = n^{-1}\{1 + \lambda^T \bar{Z}_t(\theta, \alpha)\}^{-1}$ for $t = 1, \ldots, n$ and

$$l(\theta, \alpha) := -2 \log L(\theta, \alpha) = 2 \sum_{t=1}^{n} \log\{1 + \lambda^T \bar{Z}_t(\theta, \alpha)\},$$

where $\lambda = \lambda(\theta, \alpha)$ satisfies

$$\sum_{t=1}^{n} \frac{\bar{Z}_t(\theta, \alpha)}{1 + \lambda^T \bar{Z}_t(\theta, \alpha)} = 0. \tag{5.11}$$

Since we are interested in the tail index α, we consider the profile empirical likelihood ratio $l_p(\alpha) = l(\tilde{\theta}(\alpha), \alpha)$, where $\tilde{\theta}(\alpha) = \arg\min_\theta l(\theta, \alpha)$. Next theorem shows that Wilks theorem holds for the proposed profile empirical likelihood method.

Theorem 6. *Under conditions of Theorem 5, the random variable $l_p(\alpha_0)$ converges in distribution to $\chi^2(1)$ as $n \to \infty$.*

Corollary 1. *For any* $0 < \xi < 1$*, let* $\chi^2_{1,\xi}$ *denote the* ξ*-th quantile of a* $\chi^2(1)$ *random variable and define the empirical likelihood confidence interval with level* ξ *as* $I_\xi = \{\alpha |\, l_p(\alpha) \le \chi^2_{1,\xi}\}$*. Then, under conditions of Theorem 5,* $P(\alpha_0 \in I_\xi) \longrightarrow \xi$ *as* $n \to \infty$*.*

Remark 6. *It is possible to develop similar estimation procedure and empirical likelihood method as above by replacing the LADE by other estimators such as those in Berkes and Horváth (2004 [101]).*

5.2.2 AR(1) with heavy tailed ARCH(1) noise

In this subsection, we study another time series model called the first-order autoregressive model (AR(1)) with autoregressive conditional heteroskedastic errors of order one (ARCH(1)), which is defined as

$$Y_t = a^* Y_{t-1} + \sqrt{\omega^* + b^* Y_{t-1}^2}\varepsilon_t^*, \tag{5.12}$$

where $\{\varepsilon_t^*\}$ is a sequence of independent and identically distributed random variables with zero mean and unit variance, $a^* \in \mathbb{R}$, $\omega^* > 0$ and $b^* > 0$. This model is also called a double AR model in the literature. Throughout this subsection, we assume model (5.12) satisfies the following regularity conditions:

Condition A. $\mathrm{E}\log(|a^* + \sqrt{b^*}\varepsilon_1^*|) < 0$;

Condition B. ε_t^* has a symmetric, positive and continuous Lebesgue density on $\mathbb{R}$.

Under Conditions A and B, it is known that Y_t has a heavy tail with index $\alpha > 0$, which is determined by

$$\mathrm{E}(|a^* + \sqrt{b^*}\varepsilon_t^*|^\alpha) = 1, \tag{5.13}$$

see Borkovec and Klüppelberg (2001[45]) for details. Therefore, one can estimate α by solving

$$\frac{1}{n}\sum_{t=1}^{n} |\hat{a}^* + \sqrt{\hat{b}^*}\hat{\varepsilon}_t^*|^\alpha = 1,$$

where $\hat{a}^*, \hat{b}^*, \hat{\varepsilon}_t^*$ are some estimators for $a^*, b^*, \varepsilon_t^*$, respectively. Indeed Chan, Li, Peng and Zhang (2013 [63]) proposed to first employ the QMLE in Ling (2004 [102]) to estimate α and then to apply a profile empirical likelihood method for interval estimation, where finite fourth moment of ε_t^* is required to ensure a normal limit. Here we propose to relax this moment condition by using the weighted least absolute deviations estimate in Chan and Peng (2005 [103]) as follows by observing that equation (5.13) is invariant to a scale transformation of the model.

Assume the unknown median of $(\varepsilon_t^*)^2$ is $d > 0$. Put $\varepsilon_t = \varepsilon_t^*/\sqrt{d}$. Then the median of ε_t^2 becomes one, and model (5.12) and equation (5.13) can be written as

$$Y_t = aY_{t-1} + \sqrt{\omega + bY_{t-1}^2}\varepsilon_t \tag{5.14}$$

and

$$E\{|a + \sqrt{b}\epsilon_t|^\alpha\} = 1, \tag{5.15}$$

where $a = a^*, \omega = d\omega^*$ and $b = db^*$. Therefore, as before we first propose to estimate $\theta = (\omega, a, b)^T$ by the following weighted least absolute deviations estimate

$$\hat{\theta} = (\hat{\omega}, \hat{a}, \hat{b})^T = \arg\min_{\theta} \sum_{t=1}^{n} \frac{1}{1 + Y_{t-1}^2} |(Y_t - aY_{t-1})^2 - (\omega + bY_{t-1}^2)|. \tag{5.16}$$

Put $\hat{\varepsilon}_t = (Y_t - \hat{a}Y_{t-1})/\sqrt{\hat{\omega} + \hat{b}Y_{t-1}^2}$. Then, α can be estimated by solving the following equation:

$$\frac{1}{n} \sum_{t=1}^{n} |\hat{a} + \sqrt{\hat{b}}\hat{\varepsilon}_t|^\alpha = 1. \tag{5.17}$$

Denote this estimator by $\hat{\alpha}$, and let α_0 denote the true value of α. Put $\Delta = (1+Y_1^2)(\omega_0 + b_0Y_1^2)$, $S = 1+Y_1^2$,

$$\Gamma_1 = \begin{pmatrix} \mathrm{E}\frac{a_0^2Y_1^4}{\Delta} + \mathrm{E}\frac{Y_1^2}{S} & \mathrm{E}\frac{a_0Y_1^2}{\Delta} & -\mathrm{E}\frac{a_0Y_1^4}{\Delta} \\ \mathrm{E}\frac{a_0Y_1^2}{\Delta} & \mathrm{E}\frac{1}{\Delta} & -\mathrm{E}\frac{Y_1^2}{\Delta} \\ -\mathrm{E}\frac{a_0Y_1^4}{\Delta} & -\mathrm{E}\frac{Y_1^2}{\Delta} & \mathrm{E}\frac{Y_1^4}{\Delta} \end{pmatrix}, \ \Gamma_2 = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 \\ \alpha_0 & 0 & 1 \end{pmatrix}.$$

Let $\bar{A}(t) = (Y_tY_{t-1}, 1, -Y_{t-1}^2)^T\mathrm{sgn}(\varepsilon_t^2 - 1)/(1+Y_{t-1}^2)$, $f(x)$ denote the density of ε_1,

$$\begin{aligned} \bar{\gamma}_{\alpha_0}^2 \ &= \ \{f(1)\}^{-2}(c_1,c_2,c_3)\Gamma_2\Gamma_1^{-1}\mathrm{Cov}\{\bar{A}(1)\}\Gamma_1^{-1}\Gamma_2(c_1,c_2,c_3)^T + \kappa_0^{-2}\mathrm{Var}(|a_0+\sqrt{b_0}\varepsilon_1|^{\alpha_0}) \\ &\quad -2\{f(1)\}^{-1}\kappa_0^{-1}(c_1,c_2,c_3)\Gamma_2\Gamma_1^{-1}\mathrm{E}\{\bar{A}(1)(|a_0+\sqrt{b_0}\varepsilon_1|^{\alpha_0} - \mathrm{E}|a_0+\sqrt{b_0}\varepsilon_1|^{\alpha_0})\}, \end{aligned}$$

where

$$\begin{cases} \kappa_0 = \mathrm{E}\{|a_0+\sqrt{b_0}\varepsilon_1|^{\alpha_0}\log|a_0+\sqrt{b_0}\varepsilon_1|\}, \\ c_1 = \kappa_0^{-1}\mathrm{E}\frac{\sqrt{b_0}(\alpha_0|a_0+\sqrt{b_0}\varepsilon_2|^{\alpha_0-1}\mathrm{sgn}(a_0+\sqrt{b_0}\varepsilon_2))\epsilon_2}{2(w_0+b_0Y_1^2)}, \\ c_2 = \kappa_0^{-1}\mathrm{E}\{(\alpha_0|a_0+\sqrt{b_0}\varepsilon_2|^{\alpha_0-1}\mathrm{sgn}(a_0+\sqrt{b_0}\varepsilon_2))(\frac{\sqrt{b_0}Y_1}{\sqrt{w_0+b_0Y_1^2}} - 1)\}, \\ c_3 = \kappa_0^{-1}\mathrm{E}\{(\alpha_0|a_0+\sqrt{b_0}\varepsilon_2|^{\alpha_0-1}\mathrm{sgn}(a_0+\sqrt{b_0}\varepsilon_2))(\frac{\sqrt{b_0}\varepsilon_2Y_1^2}{2(w_0+b_0Y_1^2)} - \frac{\varepsilon_2}{2\sqrt{b_0}})\}. \end{cases}$$

Then the following theorem holds.

Theorem 7. *In addition to Conditions A and B for model (5.12), we further assume that* $\alpha_0 > 1$ *and* $\mathrm{E}|\varepsilon_t|^{\delta_0} < \infty$ *for some* $\delta_0 > 2\alpha_0$. *Then as* $n \to \infty$

$$\sqrt{n}(\hat{\alpha} - \alpha_0) \xrightarrow{d} \mathrm{N}(0, \bar{\gamma}_{\alpha_0}^2).$$

Again, to avoid estimating $\bar{\gamma}_{\gamma_0}^2$, we develop a profile empirical likelihood method for constructing a confidence interval for α_0. Put $\varepsilon_t^2(\omega,a,b) = [(Y_t - aY_{t-1})^2 - (\omega + bY_{t-1}^2)]/(1+$

Y_{t-1}^2), define

$$X_{t,1}(\theta,\alpha) = \left|a + \sqrt{b}(Y_t - aY_{t-1})/\sqrt{\omega + bY_{t-1}^2}\right|^{\alpha} - 1,$$
$$X_{t,2}(\theta) = (\partial(\varepsilon_t^2(\omega,a,b))/\partial\omega)\mathrm{sgn}\{\varepsilon_t^2(\omega,a,b)\},$$
$$X_{t,3}(\theta) = (\partial(\varepsilon_t^2(\omega,a,b))/\partial a)\mathrm{sgn}\{\varepsilon_t^2(\omega,a,b)\},$$
$$X_{t,4}(\theta) = (\partial(\varepsilon_t^2(\omega,a,b))/\partial b)\mathrm{sgn}\{\varepsilon_t^2(\omega,a,b)\},$$

and write $X_t(\theta,\alpha) = (X_{t,1}(\theta,\alpha), X_{t,2}(\theta), X_{t,3}(\theta), X_{t,4}(\theta))^T$. Based on the estimating equations $\sum_{t=1}^n X_t(\theta,\alpha) = 0$, we define the empirical likelihood function of (θ,α) as

$$L(\theta,\alpha) = \sup\{\prod_{t=1}^n (np_t) : p_1 \geq 0, \ldots, p_n \geq 0, \sum_{t=1}^n p_t = 1, \sum_{t=1}^n p_t X_t(\theta,\alpha) = 0\}.$$

Put $l(\theta,\alpha) = -2\log L(\theta,\alpha)$. Since we are interested in α, we consider the profile empirical likelihood ratio $l_p(\alpha) = l(\tilde{\theta}(\alpha),\alpha)$, where $\tilde{\theta} = \tilde{\theta}(\alpha) := \arg\min_\theta l(\theta,\alpha)$. Next theorem shows that Wilks theorem holds for the proposed profile empirical likelihood method.

Theorem 8. *Under conditions of Theorem 7, $l_p(\alpha_0)$ converges in distribution to $\chi^2(1)$ as $n \longrightarrow \infty$.*

Remark 7. *Based on Theorem (8), one can construct a confidence interval for the tail index α_0 under model (5.12) as in Corollary 1.*

5.3 Data Analysis and Simulation Study

5.3.1 Data Analysis

We revisit the analysis of the daily HKD/USD exchange rate from January 21, 1998 to June 6, 2000 in Zhu and Ling (2015 [65]), where LADE-based inference is proposed to replace QMLE due to the lack of moments. Therefore it is useful to accurately estimate the tail index of this data set.

As in Zhu and Ling (2015 [65]), we consider the log-returns ($\times 100$) of this data sample denoted by $\{X_t\}_{t=1}^{600}$. First we fit an ARMA(10,10) model to the data and use the function 'auto.arima' in the R package 'forecast' with AIC to obtain the following best model:

$$\begin{array}{rcllllll} X_t & = & 0.0012+ & 0.2374X_{t-1}+ & 0.0127X_{t-2}- & 0.1536X_{t-3}- & 0.1516X_{t-4} & \\ & & (0.0004) & (0.4867) & (0.7874) & (0.2930) & (0.0689) & \\ & + & 0.0283X_{t-5}- & 1.5400e_{t-1}+ & 0.2160e_{t-2}+ & 0.3375e_{t-3}+ & e_t. & \\ & & (0.0620) & (0.4868) & (0.9383) & (0.4560) & & \end{array} \quad (5.1)$$

Denote the resulted residuals by $Y_t's$. Next we use the function 'garchFit' in the R package 'fGarch' to fit a GARCH(1,1) to $Y_t's$ and obtain

$$\begin{array}{llll} Y_t = \sigma_t\varepsilon_t, \quad \sigma_t^2 = & 1.098\times 10^{-5}+ & 0.7220\sigma_{t-1}^2+ & 0.3773Y_{t-1}^2. \\ & (9.794\times 10^{-6}) & (0.04657) & (0.14012) \end{array} \quad (5.2)$$

Numbers in brackets mean standard deviations. After this fitting, we plot $\{X_t\}$, the autocorrelation functions of $\{Y_t\}$, $\{Y_t^2\}$ and estimated $\{\varepsilon_t\}$ in Figure 1, which indicate the fitting is good. However, as showed in Zhang and Ling (2015), the estimators in (5.1) would be inconsistent theoretically if $EY_t^2=\infty$, and the standard deviations in (5.1) may be theoretically incorrect when $EY_t^4=\infty$ since this case implies that the joint asymptotic limit of estimators in (5.1) is nonnormal.

Here we study the tail index of Y_t by applying the profile empirical likelihood methods based on both the LADE in this paper and the QMLE in Chan, Peng and Zhang (2012 [62]) to $\{Y_t\}$ without taking into account of the randomness in obtaining $\{Y_t\}$. Since the parameters in (5.2) satisfy $w>0, a\in(0,1), b>0$, we rewrite $w=\exp(\tilde{w})$, $a=\exp(\tilde{a})/\{1+\exp(\tilde{a})\}$, $b=\exp(\tilde{b})$ in computing the profile empirical likelihood ratio based on QMLE. Similar transformation for θ^* in (5.1) is applied to computing the profile empirical likelihood ratio based on LADE.

First we use the R function 'garchFit' to obtain the QMLE for $\tilde{\theta}^*$, and then get an estimator for $(\epsilon_t^*)^2$, which results in an estimator for d. Hence we have initial values for both $\tilde{\theta}$ and $\tilde{\theta}^*$, which are the transformed θ and θ^*. Denote them by $\tilde{\theta}_{ini}$ and $\tilde{\theta}^*_{ini}$. Using the obtained initial value $\tilde{\theta}_{ini}$, we minimize $\Delta(\theta)=\sum_{j=1}^4(\frac{1}{n}\sum_{t=1}^n\bar{Z}_{t,j}(\theta))^2$ to obtain $\bar{\theta}_{ini}$.

Next we employ the R package 'emplik' and the R function 'optim' to compute the profile empirical likelihood ratio based on LADE for $\alpha = 0.7, 0.75, 0.8, ..., 4$ by using either $\tilde{\theta}_{ini}$ or $\bar{\theta}_{ini}$, depending on which one gives a smaller value of $\Delta(\theta)$, as an initial value. The same approach is applied to calculating the profile empirical likelihood ratio based on QMLE by using $\tilde{\theta}^*_{ini}$ instead of $\tilde{\theta}_{ini}$. We also compute the profile empirical likelihood ratios by restricting $|\tilde{\theta} - \tilde{\theta}_{ini}| \leq \delta$ and $|\tilde{\theta}^* - \tilde{\theta}^*_{ini}| \leq \delta$ with $\delta = 0.5$. Hence we use $\delta = \infty$ to mean no such a restriction in our calculation. The profile empirical likelihood ratio based on LADE in Figure 2 has its minimum around $\alpha = 1.8$ and indicates $\alpha_0 \in (1, 3)$ at both level 90% and level 95%. The profile empirical likelihood ratio based on QMLE in Figure 2 gives a very large value when $\alpha = 2$, which rejects $H_0 : \alpha_0 = 2$. This is in line with the fact that the estimated value of $a + b$ in (5.2) is larger than one, i.e., $EY_t^2 = \infty$. However, the empirical likelihood ratio based on QMLE fails to reject other considered α's in $(0.7, 4)$ at levels 90% and 95%, which may indicate the method is not applicable to this data set. After plotting the Hill's estimator in (1.9) for both $\{Y_t\}$ and estimated $\{\varepsilon_t\}$ in Figure 3, we conclude that the method in Chan, Peng and Zhang (2012 [62]) is problematic since $E\varepsilon_t^4$ seems infinite, and the standard deviations in (5.1) are inaccurate since $EY_t^4 = \infty$. Note that the 95% confidence intervals in Figure 3 are based on $\sqrt{k}(\tilde{\alpha}(k)/\alpha_0 - 1) \xrightarrow{d} N(0, 1)$ for independent data.

5.3.2 Simulation Study

In this section we examine the finite sample behavior of the proposed profile empirical likelihood for a GARCH(1,1) sequence and compare it with the method in Chan, Peng and Zhang (2012 [62]), where the errors are required to have a finite fourth moment.

Consider model (5.1) with $\varepsilon_t^* \sim t(\nu)/\sqrt{\nu/(\nu - 2)}$ with $\nu = 3.2$, or 4, or 8, or 12, and $\theta_0^* = (1, 0.72, 0.38)^T$, or $(1, 0.65, 0.38)^T$, or $(1, 0.65, 0.25)^T$, or $(1, 0.6, 0.25)^T$. By drawing $5,000$ random samples with sample size $n = 500$, $n = 1,000$ and $n = 2,000$, we follow the procedure in the data analysis to compute the profile empirical likelihood ratios and calculate the coverage probabilities for the proposed profile empirical likelihood confidence interval in this paper and that in Chan, Peng and Zhang (2012 [62]) with levels $\xi = 0.9$ and $\xi = 0.95$, which are denoted by I_ξ^{LADE} and I_ξ^{QMLE}.

Coverage probabilities for these two methods are reported in Table 1, which shows that the proposed profile empirical likelihood method works well and even performs better than the method based on the QMLE in Chan, Peng and Zhang (2012 [62]). Results for $\nu = 3$ and 4 well indicate that the method in Chan, Peng and Zhang (2012 [62]) does not work since the errors have an infinite fourth moment.

5.4 Conclusions

It is known that the tail index of a GARCH(1,1) sequence or an AR(1) model with ARCH(1) errors is determined by an estimating equation, which can be employed to estimate the tail index at the rate of $\sqrt{n}$, where n is the sample size. That is, the resulted tail index estimator has a faster rate of convergence than an estimator based on extreme value theory. However, this estimation procedure requires that the plug-in estimators for the unknown parameters in the model should have a joint normal limit, which generally needs a finite fourth moment for the errors. By noting that the estimating equation for determining the tail index is invariant to a scale transformation of the underlying model, we propose to estimate the tail index by employing some least absolute deviations estimate so as to relax the moment condition on errors. Although the resulted tail index estimator has a $\sqrt{n}$ rate of convergence and a normal limit, the asymptotic variance is quite complicated. To effectively construct a confidence interval for the tail index, we further propose a profile empirical likelihood method, which does not need to estimate any additional quantities such as asymptotic variance. A simulation study confirms that the proposed new methods have good finite sample behavior.

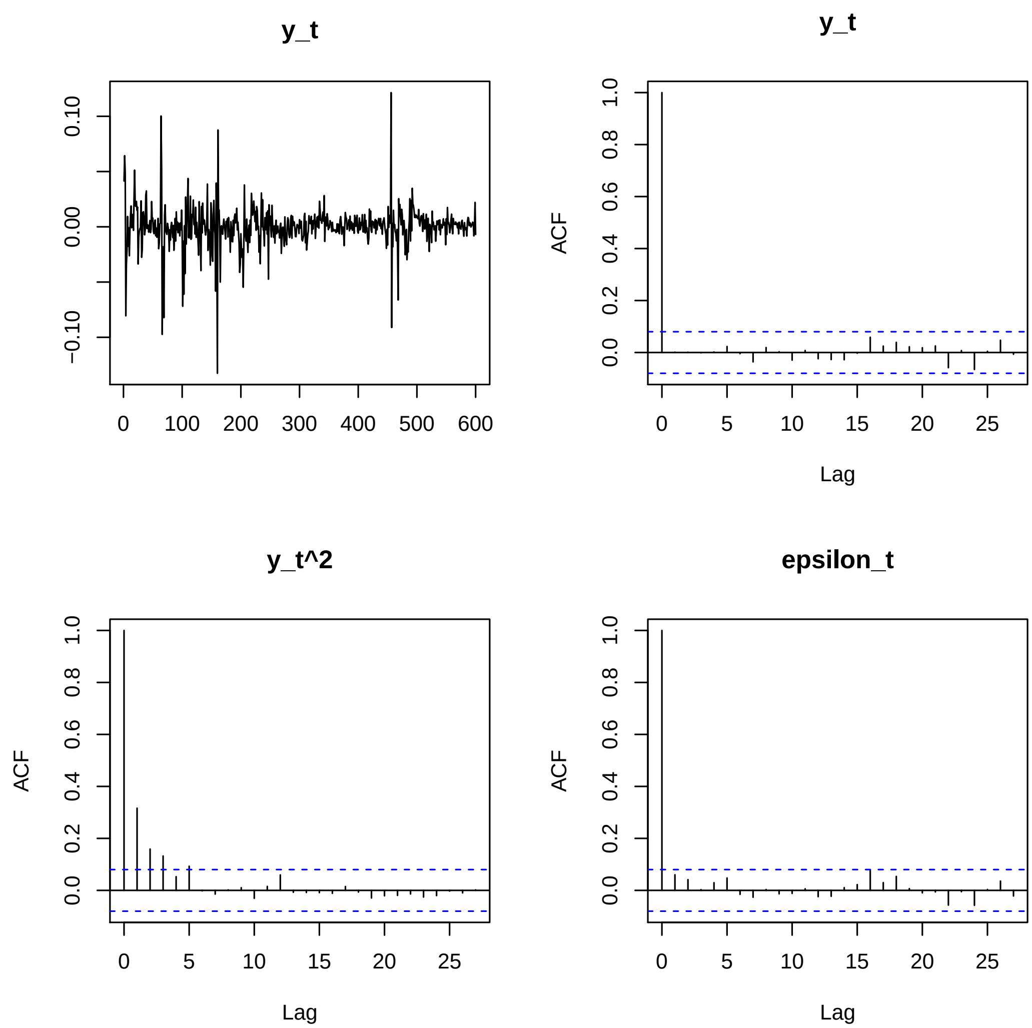


Figure (5.1) We plot $\{X_t\}$, the autocorrelation functions of $\{Y_t\}$, $\{Y_t^2\}$ and estimated $\{\varepsilon_t\}$ respectively.

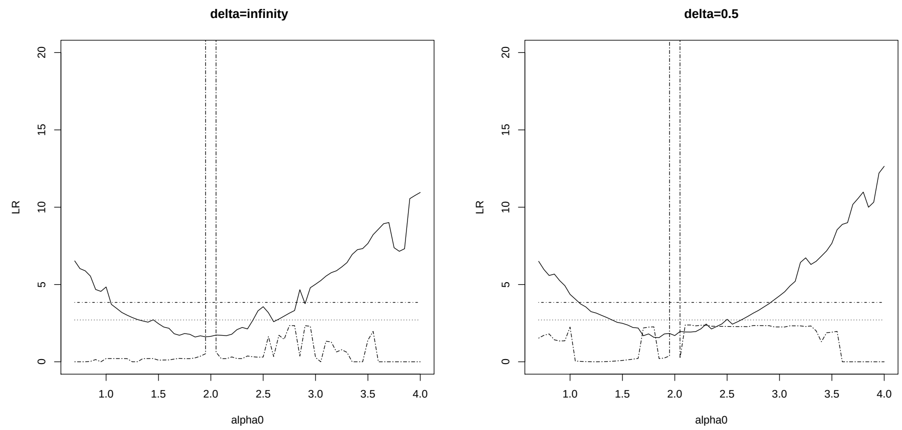


Figure (5.2) The profile empirical likelihood ratios based on both LADE and QMLE are plotted against $\alpha = 0.7, 0.75, \cdots, 4$ in solid line and dotted line, respectively. Two straight lines represent the 90th and 95th quantile of $\chi^2(1)$ respectively.

Table (5.1) Coverage probabilities based on the method in Section 2.1 (I_ξ^{LADE}) and the method in Chan, Peng and Zhang (2012) (I_ξ^{QMLE}) are calculated for $w_0 = 1$ and $\varepsilon_t^* \sim t(\nu)/\sqrt{\nu/(\nu-2)}$. Here $\nu = \infty$ means $\varepsilon_t^* \sim N(0,1)$.

(a_0, b_0, ν, n)	$I_{0.90}^{LADE}$	$I_{0.95}^{LADE}$	$I_{0.90}^{QMLE}$	$I_{0.95}^{QMLE}$	α_0
$(0.72, 0.38, 3.2, 500)$	0.8892	0.9340	0.9700	0.9882	1.1678
$(0.65, 0.38, 4, 500)$	0.8988	0.9528	0.9374	0.9682	1.7367
$(0.65, 0.25, 8, 500)$	0.9044	0.9530	0.9536	0.9734	3.8212
$(0.6, 0.25, 12, 500)$	0.9000	0.9526	0.9174	0.9546	4.7681
$(0.72, 0.38, 3.2, 1000)$	0.9052	0.9484	0.9842	0.9930	1.1678
$(0.65, 0.38, 4, 1000)$	0.9034	0.9538	0.9552	0.9848	1.7367
$(0.65, 0.25, 8, 1000)$	0.8940	0.9468	0.9504	0.9752	3.8212
$(0.6, 0.25, 12, 1000)$	0.8988	0.9514	0.9102	0.9586	4.7681
$(0.72, 0.38, 3.2, 2000)$	0.9064	0.9524	0.9854	0.9948	1.1678
$(0.65, 0.38, 4, 2000)$	0.9130	0.9616	0.9694	0.9872	1.7367
$(0.65, 0.25, 8, 2000)$	0.8978	0.9492	0.9222	0.9724	3.8212
$(0.6, 0.25, 12, 2000)$	0.9020	0.9476	0.8768	0.9416	4.7681

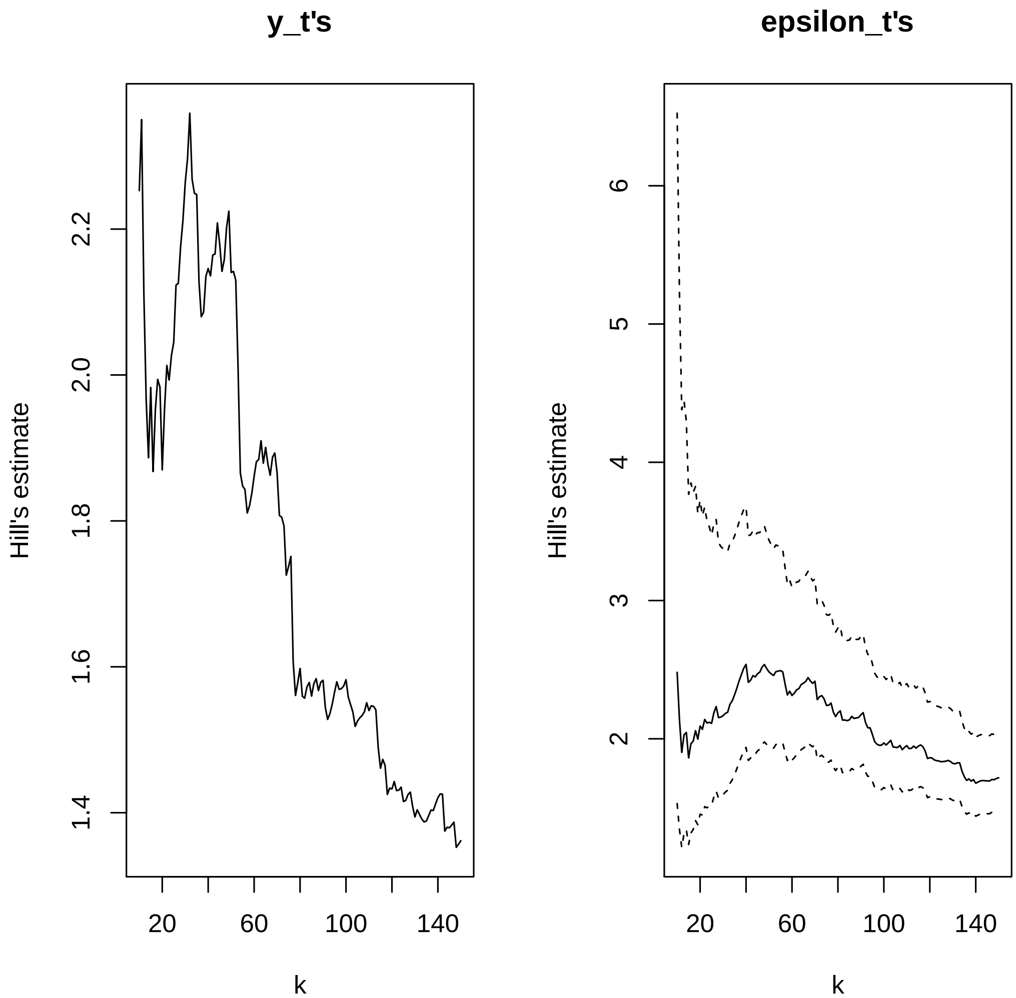


Figure (5.3) Hill's estimator in (1.9) is plotted against $k = 10, 11, \cdots, 150$ for $\{Y_t\}$ and estimated $\{\varepsilon_t\}$.

REFERENCES

[1] C. Z. Zhu, F. and L. Peng, “Predictive regressions for macroeconomic data.” *Ann. Appl. Statist.*, vol. 8, pp. 577–594, 2014.

[2] J. Bruin. (2011, FEB) Introduction to sas. ucla: Statistical consulting group. [Online]. Available: http://www.ats.ucla.edu/stat/sas/examples/alr2/hlch5sas.htm

[3] C. Pierce, *The numerical measure of the success of predictions.* Science, 1884.

[4] D. Bamber, “The area above the ordinal dominance graph and the area below the receiver operating characteristic graph,” *Journal of mathematical psychology*, vol. 12, no. 4, pp. 387–415, 1975.

[5] J. H. Shapiro, “Bounds on the area under the roc curve,” *Journal of the Optical Society of America*, vol. 16, no. 1, pp. 53–57, 1999.

[6] D. D. Zhang, X. H. Zhou, D. H. J. Freeman, and J. L. Freeman, “A non-parametric method for the comparison of partial areas under roc curves and its application to large health care data sets,” *Statist. Med*, vol. 21, pp. 701–715, 2002.

[7] M. S. Pepe, *The statistical evaluation of medical tests for classification and prediction*, Mar 2003.

[8] S. D. Walter, “The partial area under the summary roc curve,” *Statist. Med.*, no. 24, pp. 2025–2040, 2005.

[9] D. K. McClish, “Analyzing a portion of the roc curve,” *Medical Decision Making*, vol. 9, pp. 190–195, 1989.

[10] ——, “Determining a range of false positive rates for which roc curves,” *Medical Decision Making*, vol. 10, pp. 283–287, 1990.

[11] Y. Jiang, C. E. Metz, and R. M. Nishikawa, "A receiver operating characteristic partial area index for highly sensitive diagnostic tests," *Radiology*, vol. 201(3), pp. 745–750, Dec 1996.

[12] L. E. Dodd and M. S. Pepe, "Partial auc estimation and regression," *Biometrics*, no. 59, pp. 614–623, Sep 2003.

[13] R. Miller and D. Siegmund, "Maximally selected chi square statistics," *Biometrics*, no. 38, pp. 1011–1016, 1982.

[14] D. G. Altman, B. Lausen, W. Sauerbrei, and M. Schumacher, "Dangers of using 'optimal' cutpoints in the evaluation of prognostic factors," *Journal of the National Cancer Institute*, vol. 86, no. 11, pp. 829–835, 1994.

[15] W. Youden, "Index for rating diagnostic tests," *Cancer*, vol. 3, pp. 32–35, 1950.

[16] E. F. Schisterman and N. Perkins, "Confidence interval for youden index and corresponding optimal cut-point," *Communication in Statistics-Simulation and Computation*, vol. 36, pp. 549–563, 2007.

[17] R. Fluss, D. Faraggi, and B. Reiser, "Estimation of the youden index and its associated cutoff point," *Biometrical Journal*, vol. 47, pp. 458–472, 2005.

[18] F. Hsieh and W. Turnbull, "Nonparametric methods for evaluating diagnostic tests," vol. 6, pp. 47–62, 1996.

[19] D. Faraggi, "Adjusting receiver operating curves and related indices for covariates," vol. 52, no. 179-192, 2003.

[20] H. C. Zhou, "Statistical inferences for the youden index," Dec 2011.

[21] A. Agresti and B. A. Coull, "Approximate is better than 'exact' for interval estimation of binomial proportions," vol. 27, pp. 119–126, 1998.

[22] S. Wieand, M. H. Gail, B. R. James, and K. L. James, "A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data," *Biometrika*, vol. 76, no. 3, pp. 585–592, 1989.

[23] P. Smith and T. Thompson, "Correcting for confounding in analyzing receiver operating characteristic curves," *Biometrical Journal*, vol. 38, no. 7, pp. 857–863, 1996.

[24] L. E. Dodd, "Regression methods for areas and partial areas under the receiver-operating characteristic curve," *Thesis (Ph. D.)–University of Washington*, 2001.

[25] M. S. Pepe, "Three approaches to regression analysis of receiver operating characteristic curves for continuous test results," *Biometrics*, pp. 124–135, 1998.

[26] A. N. A. Tosteson and C. B. Begg, "A general regression methodology for roc curve estimation," *Medical Decision Making*, vol. 8, no. 3, pp. 204–215, 1988.

[27] X.-H. Zhou, D. K. McClish, and N. A. Obuchowski, *Statistical methods in diagnostic medicine*. John Wiley & Sons, 2009, vol. 569.

[28] Y. Zheng and P. J. Heagerty, "Semiparametric estimation of time-dependent roc curves for longitudinal marker data," *Biostatistics*, vol. 5, no. 4, pp. 615–632, 2004.

[29] A. Rodríguez and J. C. Martínez, "Bayesian semiparametric estimation of covariate-dependent roc curves," *Biostatistics*, vol. 15, no. 2, pp. 353–369, 2014.

[30] D. Faraggi, "Adjusting receiver operating characteristic curves and related indices for covariates," *Journal of the Royal Statistical Society: Series D (the Statistician)*, vol. 52, no. 2, pp. 179–192, 2003.

[31] R. Stambaugh, "Predictive regressions." *Journal of Financial Economics*, vol. 54, pp. 375–421, 1999.

[32] Y. Amihud and C. Hurvich, "Predictive regressions: a reduced-bias estimation method." *J. Financ. Quant. Anal.*, vol. 39, pp. 813–841, 2004.

[33] H. C. Amihud, Y. and Y. Wang, "Multiple-predictor regressions: hypothesis testing." *Rev. Financ. Stud.*, vol. 22, pp. 413–434, 2009.

[34] W. Chen and R. Deo, "Bias reduction and likelihood-based almost exactly sized hypothesis testing in predictive regressions using the restricted likelihood." *Econometric Theory*, vol. 25, pp. 1143–1179, 2009.

[35] E. G. Cavanagh, C.L. and J. Stock, "Inference in models with nearly integrated regressors." *Econometric Theory*, vol. 11, pp. 1131–1147, 1995.

[36] J. Campbell and M. Yogo, "Efficient tests of stock return predictability." *Journal of Financial Economics*, vol. 81, pp. 27–60, 2006.

[37] S. J. Taylor, *Asset Price Dynamics, Volatility, and Prediction*, N. Jersey, Ed. Princeton University Press, 2005.

[38] R. A. Davis and S. I. Resnick, "Limit theory for moving averages of random variables with regularly varying tail probabilities." *Ann. Probab.*, vol. 13, pp. 179–195, 1985.

[39] ——, "Limit theory for the sample covariance and correlation functions of moving averages." *Ann. Statist.*, vol. 14, pp. 533–558, 1986.

[40] T. Mikosch and C. Stărică, "Limit theory for the sample autocorrelations and extremes of a garch (1,1) process." *Ann. Statist.*, vol. 28, pp. 1427–1451, 2000.

[41] D. R. A. Basrak, B. and T. Mikosch, "Regular variation of garch processes." *Stochastic Processes and Their Applications*, vol. 99, pp. 95–115, 2002.

[42] R. A. Davis and T. Mikosch, "The sample autocorrelations of heavy-tailed processes with applications to arch." *Ann. Statist.*, vol. 26, pp. 2049–2080, 1998.

[43] R. S. I. R. H. De Haan, L. and C. G. de Vries, "Extremal behavior of solutions to a stochastic difference equation with applications to arch processes." *Stochastic Processes and Their Applications*, vol. 32, pp. 213–224, 1989.

[44] M. Borkovec, “Extremal behavior of the autoregressive process with arch(1) errors.” *Stochastic Processes and Their Applications*, 2000.

[45] ——, “Asymptotic behavior of the sample autocovariance and autocorrelation function of the ar(1) process with arch errors.” *Bernoulli*, vol. 7, pp. 847–872, 2001.

[46] R. A. Davis and S. I. Resnick, “Limit theory for bilinear processes with heavy tailed noise.” *Ann. Appl. Probab.*, vol. 6, pp. 1191–1210, 1996.

[47] S. I. Resnick and E. Van den Berg, “Sample correlation behavior for the heavy tailed general bilinear process.” *Stochastic Models*, vol. 16, pp. 233–258, 2000.

[48] P. Hall and Q. Yao, “Inference in arch and garch models.” *Econometrica*, vol. 71, pp. 285–317, 2003.

[49] T. Lange, “Tail behavior and ols estimation in ar-garch models.” *Statistica Sinica*, vol. 21, pp. 1191–1200, 2011.

[50] R. M. Zhang and S. Ling, “Asymptotic inference for ar models with heavy-tailed g-garch noises.” *Econometric Theory*, vol. 31, pp. 880–890, 2015.

[51] J. Hill, “Robust estimation and inference for heavy tailed garch.” *Bernoulli*, vol. 21, pp. 1629–1669, 2015.

[52] J. Hill and A. Prokhorov, “Gel estimation for heavy-tailed garch models with robust empirical likelihood inference.” *J. Econometrics*, vol. 190, pp. 18–45, 2016.

[53] F. Z. Quintos, C. and P. C. B. Phillips, “Structural change tests in tail behaviour and the asian crisis.” *Review of Economic Studies*, vol. 68, pp. 633–663, 2001.

[54] N. Wagner and T. A. Marsh, “Measuring tail thickness under garch and an application to extreme exchange rate changes.” *Journal of Empirical Finance*, vol. 12, pp. 165–185, 2005.

[55] H. Kesten, "Random difference equations and renewal theory for products of random matrices." *Acta Math*, vol. 131, pp. 207–248, 1973.

[56] T. Hsing, "On tail index estimation using dependent data." *Ann. Statist.*, vol. 19, pp. 1547–1569, 1991.

[57] S. I. Resnick and C. Starica, "Tail index estimation for dependent data." *Ann. Appl. Probab.*, vol. 8, pp. 1156–1183, 1998.

[58] H. Drees, "Weighted approximations for tail processes for β-mixing random variables." *Ann. Appl. Probab.*, vol. 10, pp. 1274–1301, 2000.

[59] S. I. Resnick and C. Starica, "Asymptotic behavior of hill's estimator for autoregressive data." *Stochastic Models*, vol. 13, pp. 703–723, 1997.

[60] S. Ling and L. Peng, "Hill's estimator for the tail index of an arma model." *Journal of Statistical Planning and Inference*, vol. 123, pp. 279–293, 2004.

[61] H. L. Berkes, I. and P. Kokoszka, "Estimation of the maximal moment exponent of a garch (1,1) sequence." *Econometric Theory*, vol. 19, pp. 565–586, 2003.

[62] P. L. Chan, N. H. and R. M. Zhang, "Interval estimation for garch(1,1) tail index." *Test*, vol. 21, pp. 546–565, 2012.

[63] L. D. Y. P. L. Chan, N. H. and R. M. Zhang, "Interval estimation of the tail index of an ar(1) with arch(1) errors." *Econometric Theory*, 2013.

[64] Z. Cai and Y. Wang, "Testing predictive regression models with nonstationary regressors." *J. Econometrics*, vol. 178, pp. 4–14, 2014.

[65] K. Zhu and S. Ling, "Lade-based inference for arma models with unspecified and heavy-tailed heteroscedastic noises." *Journal of American Statistical Association*, vol. 110, pp. 784–794, 2015.

[66] L. D. Chan, N.H. and L. Peng, "Toward a unified interval estimation of autoregressions." *Econometric Th.*, vol. 28, pp. 705–717, 2012.

[67] S. Weerahandi, "Generalized confidence intervals," vol. 88, pp. 899–905, 1993.

[68] G. Zou and A. Donner, "Construction of confidence limits about effect measures: A general approach," *Statistics in Medicine*, no. 27, pp. 1693–1702, 2008.

[69] M. L. Thompson and W. Zucchini, "On the statistical analysis of roc curves," *Statistics in Medicine*, vol. 8, pp. 1277–1290, 1989.

[70] R. Newcombe, "Interval estimation for the difference between independent proportions: Comparison of eleven methods," *Statistics in Medicine*, no. 17, pp. 873–890, 1998.

[71] T. J. H. C. Zou, G., "Confidence interval estimation for lognormal data with application to health economic," *Computational Statistics & Data Analysis*, no. 53, pp. 3755–3764, 2009.

[72] G. Shan, "Improved confidence intervals for the youden index," *PLoS ONE*, vol. 10, no. 7, July 2015.

[73] X. Huang, G. S. Qin, Y. Yuan, and X. H. Zhou, "Confidence intervals for the difference between two partial aucs," *Australian & New Zealand Journal of Statistics*, vol. 54, no. 1, pp. 63–79, 2012.

[74] W. S. Tsui KW, "Generalized p-values insignificance testing of hypotheses in the p-resence of nuisance parameters," *Journal of American Statistical Association*, no. 84, pp. 602–307, 1989.

[75] J. Gamage, T. Mathew, and S. Weerahandi, "Generalized p-values and generalized confidence regions for the multivariate behrens–fisher problem and manova," *Journal of Multivariate Analysis*, vol. 88, no. 1, pp. 177–189, 2004.

[76] J. C. Lee and S.-H. Lin, "Generalized confidence intervals for the ratio of means of two normal populations," *Journal of statistical planning and inference*, vol. 123, no. 1, pp. 49–60, 2004.

[77] L. Tian and G. E. Wilding, "Confidence interval estimation of a common correlation coefficient," *Computational Statistics & Data Analysis*, vol. 52, no. 10, pp. 4872–4877, 2008.

[78] C.-Y. Lai, L. Tian, and E. F. Schisterman, "Exact confidence interval estimation for the youden index and its corresponding optimal cut-point," *Computational Statistics and Data Analysis*, vol. 56, no. 5, pp. 1103–1114, 2012.

[79] S. Weerahandi, *Generalized inference in repeated measures: exact methods in MANOVA and mixed models.* John Wiley & Sons, 2004, vol. 500.

[80] ——, *Exact statistical methods for data analysis.* Springer Science & Business Media, 2013.

[81] E. F. Schisterman and N. Perkins, "Confidence intervals for the youden index and corresponding optimal cut-point," *Communications in Statistics–Simulation and Computation*, vol. 36, no. 3, pp. 549–563, 2007.

[82] Q. Ibrahim, "Generalized inference in linear regression models," *University of Windsor Scholarship at UWindsor*, 2009.

[83] W. González-Manteiga, J. C. Pardo-Fernández, and I. V. KEILEGOM, "Roc curves in non-parametric location-scale regression models," *Scandinavian Journal of Statistics*, vol. 38, no. 1, pp. 169–184, 2011.

[84] V. I. de Carvalho, A. Jara, T. E. Hanson, M. de Carvalho *et al.*, "Bayesian nonparametric roc regression modeling," *Bayesian Analysis*, vol. 8, no. 3, pp. 623–646, 2013.

[85] J. C. Pardo-Fernández, M. X. Rodríguez-Álvarez, I. Van Keilegom *et al.*, "A review on roc curves in the presence of covariates," UCL, Tech. Rep., 2013.

[86] C.-R. Li, C.-T. Liao, and J.-P. Liu, "On the exact interval estimation for the difference in paired areas under the roc curves," *Statistics in medicine*, vol. 27, no. 2, pp. 224–242, 2008.

[87] L. C.-Y. V. A. Tian L, Xiong C, "Exact confidence interval estimation for the difference in diagnostic accuracy with three ordinal diagnostic groups," *Journal of Statistical Planning and Inference*, no. 141, pp. 549–558, 2011.

[88] L. O. C. R. Xiao, Z. and E. Mammen, "More efficient local polynomial estimation in nonparametric regression with autocorrelated errors." *J. Amer. Statist. Assoc.*, vol. 98, pp. 980–992, 2003.

[89] C. R. Liu, J.M. and Q. Yao, "Nonparametric transfer function models." *Journal of Econometircs*, vol. 157, pp. 151–164, 2010.

[90] P. Hall and Q. Yao, "Date tilting for time series." *J. Royal Statist. Soc. B.*, vol. 65, pp. 425–442, 2003.

[91] J. Hill and L. Peng, "Unified interval estimation for random coefficient autoregressive models." *J. Time Ser. Anal.*, vol. 35, pp. 282–297, 2014.

[92] J. Qin and J. Lawless, "Empirical likelihood and general estimating functions." *Ann. Statist.*, vol. 22, pp. 300–325, 1994.

[93] C. N. Li, D. and L. Peng, "Empirical likelihood test for causality for bivariate ar(1) processes." *Econometric Th.*, vol. 30, pp. 357–371, 2014.

[94] Y. J. Jing, B.Y. and W. Zhou, "Jackknife empirical likelihood." *J. Amer. Statist. Assoc.*, vol. 104, pp. 1224–1232, 2009.

[95] P. L. Li, M. and Y. Qi, "Reduce computation in profile empirical likelihood method." *Can. J. Stat.*, vol. 39, pp. 370–384, 2011.

[96] C. Francq and J. Zakoïan, *GARCH Models: Structure, Statistical Inference and Financial Applications.*, C. Francq and J. Zakoïan, Eds. John Wiley & Sons, 2010.

[97] C. M. Goldie, "Implicit renewal theory and tails of solutions of random equations." *Ann. Appl. Probab.*, vol. 1, pp. 126–166, 1991.

[98] Q. L. Fan, J. and D. Xiu, "Quasi maximum likelihood estimation of garch models with heavy-tailed likelihoods." *Journal of Business and Economics Statistics*, vol. 32, pp. 178–205, 2014.

[99] L. Peng and Q. Yao, "Least absolute deviations estimation for arch and garch models." *Biometrika*, vol. 90, pp. 967–975, 2003.

[100] J. Fan and Q. Yao, *Nonlinear Time Series: Nonparametric and Parametric Methods.*, N. York, Ed. Springer, 2003.

[101] I. Berkes and L. Horváth, "The efficiency of the estimators of the parameters in garch processes." *Ann. Stat.*, vol. 32, pp. 633–655, 2004.

[102] S. Ling, "Estimation and testing stationarity for double autoregressive models." *J. R. Statist. Soc. B*, vol. 66, pp. 63–78, 2004.

[103] N. H. Chan and L. Peng, "Weighted least absolute deviations estimation for an ar(1) process with arch(1) errors." *Biometrika*, vol. 92, pp. 477–484, 2005.

[104] P. Hall and C. C. Heyde, *Martingale Limit Theory and Its Application.*, N. York, Ed. Academic Press, 1980.

Appendix A

PROOFS OF CHAPTER 3

Recall we use $\alpha_{0,1}, \alpha_{0,2}, \beta_0, \gamma_0 = (\gamma_{0,1}, \cdots, \gamma_{0,p})^T$ to denote the true values of $\alpha_1, \alpha_2, \beta$ and γ, respectively. Our proofs are along the lines of Hill, Li and Peng (2014). Before proving Theorem 1, we need the following lemmas. Throughout when we say $|\phi_0| < 1$, we also mean that ϕ_0 is independent of the sample size n.

Lemma 1. *Under conditions of Theorem 1, we have* $\frac{1}{\sqrt{n}} \sum_{t=1}^n Z_t(\beta_0, \gamma_0) \xrightarrow{d} N(0, \Sigma)$ *as* $n \to \infty$*, where* $\Sigma = (\sigma_{i,j})_{1 \le i,j \le p+1}$ *with* $\sigma_{i,j} = E(\varepsilon_t^2) E(U_{t-i} U_{t-j})$ *for* $i, j = 1, \cdots, p$,

$$\sigma_{i,p+1} = \begin{cases} E(\varepsilon_t^2) E\left\{U_{t-i}\left(\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} + \sum_{j=1}^p \gamma_{0,j} \frac{X_{t-j-1}}{\sqrt{1+X_{t-j-1}^2}}\right)\right\} & \textit{when} \quad |\phi_0| < 1, \\ 0 & \textit{when} \quad \phi_0 = 1 - \delta_0/n \end{cases}$$

for $i = 1, \cdots, p$ *and*

$$\sigma_{p+1,p+1} = \begin{cases} E(\varepsilon_t^2) E\left\{\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} + \sum_{j=1}^p \gamma_{0,j} \frac{X_{t-j-1}}{\sqrt{1+X_{t-j-1}^2}}\right\}^2 & \textit{when} \quad |\phi_0| < 1, \\ E(\varepsilon_t^2) \times \left(1 + \sum_{j=1}^p \gamma_{0,j}\right)^2 & \textit{when} \quad \phi_0 = 1 - \delta_0/n. \end{cases}$$

Proof. Note that

$$\begin{cases} Z_{t,j}(\beta_0, \gamma_0) & = \varepsilon_t U_{t-j} \text{ for } j = 1, \cdots, p, \\ Z_{t,p+1}(\beta_0, \gamma_0) & = \varepsilon_t \left\{\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} + \sum_{j=1}^p \gamma_{0,j} \frac{X_{t-j-1}}{\sqrt{1+X_{t-j-1}^2}}\right\}. \end{cases} \tag{A.1}$$

In the local to unity case, i.e., $\phi_0 = 1 - \delta_0/n$, we have $|X_t| \xrightarrow{p} \infty$ as $t \to \infty$, and hence

$$\left(\frac{X_{t-1}}{\sqrt{1 + X_{t-1}^2}} + \sum_{j=1}^p \gamma_{0,j} \frac{X_{t-j-1}}{\sqrt{1 + X_{t-j-1}^2}}\right)^2 \xrightarrow{p} \left(1 + \sum_{j=1}^p \gamma_{0,j}\right)^2 \text{ as } t \to \infty. \tag{A.2}$$

The lemma follows from (B.2), (B.3) and the central limit theorem for martingale differences. See Hall and Heyde (1980[104]). □

Lemma 2. *Under conditions of Theorem 1,* $n^{-1}\sum_{t=1}^{n} Z_t^T(\beta_0,\gamma_0)Z_t(\beta_0,\gamma_0) \xrightarrow{p} \Sigma$ *as* $n \to \infty$.

Proof. The claim follows instantly from (B.2)–(B.3) and the weak law of large numbers for martingale differences (see Hall and Heyde (1980[104])). □

Lemma 3. *Under conditions of Theorem 1, as* $n \to \infty$*, with probability one* $L(\beta_0,\gamma)$ *attains its maximum value at some point* $\tilde{\gamma}$ *in the interior of the ball* $||\gamma-\gamma_0|| \le n^{-d/3}$*, and* $\tilde{\gamma}$ *and* $\tilde{\lambda}$ *satisfy* $Q_{1n}(\tilde{\gamma},\tilde{\lambda}) = 0$ *and* $Q_{2n}(\tilde{\gamma},\tilde{\lambda}) = 0$*, where*

$$Q_{1n}(\gamma,\lambda) := \frac{1}{n}\sum_{i=1}^{n}\frac{Z_i(\beta_0,\gamma)}{1+\lambda^T Z_i(\beta_0,\gamma)} \quad \textit{and} \quad Q_{2n}(\gamma,\lambda) := \frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\lambda^T Z_i(\beta_0,\gamma)}\left(\frac{\partial Z_i(\beta_0,\gamma)}{\partial\gamma}\right)^T\lambda.$$

Proof. The proof is similar to the proof of Lemma 1 in Qin and Lawless (1994[92]) by using Lemmas 1 and 2. □

Proof of Theorem 1. Apply Lemmas 1–3 and arguments in Qin and Lawless (1994[92]). □

Proof of Theorem 2. The proof is similar to the proof of Theorem 1. □

Before proving Theorems 3 and 4, we need some notations and lemmas. Put

$$A = \begin{pmatrix} \frac{1}{n}\sum_{t=1}^{n}\varepsilon_t U_{t-1} \\ \cdot \\ \cdot \\ \cdot \\ \frac{1}{n}\sum_{t=1}^{n}\varepsilon_t U_{t-p} \end{pmatrix}, \quad A^{(i)} = \begin{pmatrix} \frac{1}{n-1}\sum_{t=1,t\neq i}^{n}\varepsilon_t U_{t-1} \\ \cdot \\ \cdot \\ \cdot \\ \frac{1}{n-1}\sum_{t=1,t\neq i}^{n}\varepsilon_t U_{t-p} \end{pmatrix},$$

$$B = \begin{pmatrix} \frac{1}{n}\sum_{t=1}^{n} U_{t-1}^2 & \cdots & \frac{1}{n}\sum_{t=1}^{n} U_{t-p}U_{t-1} \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ \frac{1}{n}\sum_{t=1}^{n} U_{t-p}U_{t-1} & \cdots & \frac{1}{n}\sum_{t=1}^{n} U_{t-p}^2 \end{pmatrix},$$

$$B^{(i)} = \begin{pmatrix} \frac{1}{n-1}\sum_{t=1,t\neq i}^{n} U_{t-1}^2 & \cdots & \frac{1}{n-1}\sum_{t=1,t\neq i}^{n} U_{t-p}U_{t-1} \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ \frac{1}{n-1}\sum_{t=1,t\neq i}^{n} U_{t-p}U_{t-1} & \cdots & \frac{1}{n-1}\sum_{t=1,t\neq i}^{n} U_{t-p}^2 \end{pmatrix},$$

$$\bar{\Sigma} = E(B), \quad D = (\bar{\Sigma} - B)\bar{\Sigma}^{-1}A, \quad D^{(i)} = (\bar{\Sigma} - B^{(i)})\bar{\Sigma}^{-1}A^{(i)},$$

where $i = 1, \cdots, n$.

Lemma 4. *Under the conditions of Theorem 3, we have*

$$\hat{\gamma}(\beta_0) - \gamma_0 + \bar{\Sigma}^{-1}A = O_p(n^{-1}), \tag{A.3}$$

$$\max_{1\leq i\leq n} |\hat{\gamma}^{(i)}(\beta_0) - \gamma_0 + \bar{\Sigma}^{-1}A^{(i)}| = O_p(n^{-1}) \tag{A.4}$$

and

$$\max_{1\leq i\leq n} |B(\hat{\gamma}^{(i)}(\beta_0) - \hat{\gamma}(\beta_0)) + B\bar{\Sigma}^{-1}(A^{(i)} - A) + D^{(i)} - D| = o_p(n^{-3/2}). \tag{A.5}$$

Proof. Equation (A.3) follows from

$$0 = A + B(\hat{\gamma}(\beta_0) - \gamma_0) = D + B(\hat{\gamma}(\beta_0) - \gamma_0 + \bar{\Sigma}^{-1}A), \tag{A.6}$$

$A = O_p(n^{-1/2})$ and $\bar{\Sigma} - B = O_p(n^{-1/2})$. Write

$$B^{(i)} = \frac{n}{n-1}B - \frac{B_i^*}{n-1}, \quad A^{(i)} = \frac{n}{n-1}A - \frac{A_i^*}{n-1}, \tag{A.7}$$

where

$$B_i^* = \begin{pmatrix} U_{i-1}^2 & \cdots & U_{i-p}U_{i-1} \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ U_{i-p}U_{i-1} & \cdots & U_{i-p}^2 \end{pmatrix}, \quad A_i^* = \begin{pmatrix} \varepsilon_i U_{i-1} \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_i U_{i-p} \end{pmatrix}.$$

Since $\max_{1\le i\le n, 1\le j\le p} |U_{i-1}U_{i-j}| = o_p(n^{1/2})$, it follows from (A.7) that

$$B^{(i)} = O_p(1) \quad \text{and} \quad B^{(i)} - B = o_p(n^{-1/2}) \quad \text{uniformly in} \quad i = 1, \cdots, n. \tag{A.8}$$

Similarly,

$$\bar{\Sigma} - B^{(i)} = O_p(n^{-1/2}), A^{(i)} = O_p(n^{-1/2}) \text{ and } A^{(i)} - A = o_p(n^{-1/2}) \text{ uniformly in } i = 1, \cdots, n. \tag{A.9}$$

Therefore, equation (A.4) follows from (A.9) and $0 = D^{(i)} + B^{(i)}(\hat{\gamma}^{(i)}(\beta_0) - \gamma_0 + \bar{\Sigma}^{-1}A^{(i)})$. By writing

$$0 = D^{(i)} + (B^{(i)} - B)(\hat{\gamma}^{(i)}(\beta_0) - \gamma_0 + \bar{\Sigma}^{-1}A^{(i)}) + B(\hat{\gamma}^{(i)}(\beta_0) - \gamma_0 + \bar{\Sigma}^{-1}A^{(i)}),$$

equation (A.5) follows from (A.8), (A.4) and (A.6). □

Lemma 5. *Under the conditions of Theorem 3, we have*

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} Z_{n,t}^*(\beta_0) \overset{d}{=} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \varepsilon_t \tilde{e}_t + o_p(1) \xrightarrow{d} N(0, E(\varepsilon_t^2\tilde{e}_t^2)) \quad \textit{and} \quad \frac{1}{n}\sum_{t=1}^{n} Z_{n,t}^{*2}(\beta_0) \xrightarrow{p} E(\varepsilon_t^2\tilde{e}_t^2)$$

as $n \to \infty$, *where*

$$\begin{array}{rl} \tilde{e}_t & = \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} + \sum_{k=1}^{p} \gamma_{0,k} \frac{X_{t-k-1}}{\sqrt{1+X_{t-k-1}^2}} \\ & - \begin{pmatrix} U_{t-1} \\ \cdot \\ \cdot \\ \cdot \\ U_{t-p} \end{pmatrix}^T \begin{pmatrix} E(U_p^2) & \cdots & E(U_pU_1) \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ E(U_pU_1) & \cdots & E(U_1^2) \end{pmatrix} \begin{pmatrix} E\left\{U_p\left(\frac{X_p}{\sqrt{1+X_p^2}} + \sum_{k=1}^{p} \gamma_{0,k} \frac{X_{p-k}}{\sqrt{1+X_{p-k}^2}}\right)\right\} \\ \cdot \\ \cdot \\ \cdot \\ E\left\{U_1\left(\frac{X_p}{\sqrt{1+X_p^2}} + \sum_{k=1}^{p} \gamma_{0,k} \frac{X_{p-k}}{\sqrt{1+X_{p-k}^2}}\right)\right\} \end{pmatrix}. \end{array}$$

Proof. By (A.7)–(A.9) and

$$\sum_{i=1}^{n}(A^{(i)} - A) = 0 \quad \text{and} \quad \sum_{i=1}^{n}(B^{(i)} - B) = 0, \tag{A.10}$$

we have

$$\begin{array}{rl} \sum_{i=1}^{n}\{D^{(i)} - D\} & = \sum_{i=1}^{n}(B - B^{(i)})\bar{\Sigma}^{-1}(A^{(i)} - A) \\ & = \sum_{i=1}^{n}(\frac{n}{n-1}B - B^{(i)})\bar{\Sigma}^{-1}(A^{(i)} - \frac{n}{n-1}A) - \frac{n}{(n-1)^2}B\bar{\Sigma}^{-1}A \\ & = -\frac{1}{(n-1)^2}\sum_{i=1}^{n} B_i^*\bar{\Sigma}^{-1}A_i^* + O_p(n^{-1}) \\ & = O_p(n^{-1}) \end{array} \tag{A.11}$$

and

$$\begin{array}{rl} D^{(i)} - D & = A^{(i)} - A - (B^{(i)} - B)\bar{\Sigma}^{-1}A^{(i)} - B\bar{\Sigma}^{-1}(A^{(i)} - A) \\ & = o_p(n^{-1/2}) \quad \text{uniform in} \quad i = 1, \cdots, n. \end{array} \tag{A.12}$$

Using (A.5), (A.10), (A.11) and (A.12), we can show that, for any $p \times p$ matrix Δ,

$$\begin{aligned} no_p(n^{-3/2}) &= \sum_{i=1}^{n}\{\Delta(\hat{\gamma}(\beta_0) - \hat{\gamma}^{(i)}(\beta_0)) - \Delta\bar{\Sigma}^{-1}(A^{(i)} - A) - \Delta B^{-1}(D^{(i)} - D)\} \\ &= \sum_{i=1}^{n}\Delta(\hat{\gamma}(\beta_0) - \hat{\gamma}^{(i)}(\beta_0)) + o_p(n^{-1/2}), \end{aligned}$$

$$\begin{aligned}
&\textstyle\sum_{i=1}^n (A^{(i)} - A)^T \bar{\Sigma}^{-1} \Delta \bar{\Sigma}^{-1} (A^{(i)} - A) \\
=\ &\textstyle\sum_{i=1}^n (A^{(i)} - \frac{n}{n-1} A)^T \bar{\Sigma}^{-1} \Delta \bar{\Sigma}^{-1} (A^{(i)} - \frac{n}{n-1} A) + O_p(n^{-1}) \\
=\ &\textstyle\frac{1}{(n-1)^2} \sum_{i=1}^n A_i^{*T} \bar{\Sigma}^{-1} \Delta \bar{\Sigma}^{-1} A_i^* + O_p(n^{-1}) \\
=\ &O_p(n^{-1}),
\end{aligned}$$

$$\begin{aligned}
&\textstyle\sum_{i=1}^n (A^{(i)} - A)^T \bar{\Sigma}^{-1} \Delta B^{-1} (D^{(i)} - D) \\
=\ &\textstyle\sum_{i=1}^n (A^{(i)} - \frac{n}{n-1} A)^T \bar{\Sigma}^{-1} \Delta B^{-1} (D^{(i)} - D) + O_p(n^{-2}) \\
=\ &\textstyle-\frac{1}{n-1} \sum_{i=1}^n A_i^{*T} \bar{\Sigma}^{-1} \Delta B^{-1} (D^{(i)} - D) + O_p(n^{-2}) \\
=\ &O_p(n^{-1/2}) o_p(n^{-1/2}) + O_p(n^{-2}) = O_p(n^{-1}),
\end{aligned}$$

$$\begin{aligned}
&\textstyle\sum_{i=1}^n (D^{(i)} - D)^T B^{-1} \Delta B^{-1} (D^{(i)} - D) \\
=\ &\textstyle\sum_{i=1}^n \{A^{(i)} - A\}^T B^{-1} \Delta B^{-1} (D^{(i)} - D) \\
&\textstyle- \sum_{i=1}^n \{(B^{(i)} - B) \bar{\Sigma}^{-1} A^{(i)}\}^T B^{-1} \Delta B^{-1} (D^{(i)} - D) \\
&- \{B \bar{\Sigma}^{-1} (A^{(i)} - A)\}^T B^{-1} \Delta B^{-1} (D^{(i)} - D) \\
=\ &O_p(n^{-1}) + n o_p(n^{-1/2}) O_p(n^{-1/2}) o_p(n^{-1/2}) + O_p(n^{-1}) \\
=\ &o_p(n^{-1/2}),
\end{aligned}$$

$$\begin{aligned}
&\textstyle\sum_{i=1}^n (A^{(i)} - A)^T \bar{\Sigma}^{-1} \Delta (\hat{\gamma}(\beta_0) - \hat{\gamma}^{(i)}(\beta_0)) \\
=\ &\textstyle\sum_{i=1}^n (A^{(i)} - A)^T \bar{\Sigma}^{-1} \Delta \{\hat{\gamma}(\beta_0) - \hat{\gamma}^{(i)}(\beta_0) - \bar{\Sigma}^{-1}(A^{(i)} - A) - B^{-1}(D^{(i)} - D)\} + O_p(n^{-1}) \\
=\ &n o_p(n^{-1/2}) o_p(n^{-3/2}) + O_p(n^{-1}) = O_p(n^{-1}),
\end{aligned}$$

$$\sum_{i=1}^n (D^{(i)} - D)^T B^{-1} \Delta (\hat{\gamma}(\beta_0) - \hat{\gamma}^{(i)}(\beta_0)) = O_p(n^{-1}),$$

and

$$\begin{aligned}
n o_p(n^{-3/2}) o_p(n^{-3/2}) &= \sum_{i=1}^n \left\{\hat{\gamma}(\beta_0) - \hat{\gamma}^{(i)}(\beta_0) - \bar{\Sigma}^{-1}(A^{(i)} - A) - B^{-1}(D^{(i)} - D)\right\}^T \Delta \\
&\quad \times \left\{\hat{\gamma}(\beta_0) - \hat{\gamma}^{(i)}(\beta_0) - \bar{\Sigma}^{-1}(A^{(i)} - A) - B^{-1}(D^{(i)} - D)\right\} \\
&= \sum_{i=1}^n (\hat{\gamma}(\beta_0) - \hat{\gamma}^{(i)}(\beta_0))^T \Delta (\hat{\gamma}(\beta_0) - \hat{\gamma}^{(i)}(\beta_0)) + o_p(n^{-1/2}),
\end{aligned}$$

which imply that

$$\begin{cases} \sum_{i=1}^n \Delta(\hat{\gamma}(\beta_0) - \hat{\gamma}^{(i)}(\beta_0)) = o_p(n^{-1/2}), \\ \sum_{i=1}^n (\hat{\gamma}(\beta_0) - \hat{\gamma}^{(i)}(\beta_0))^T \Delta(\hat{\gamma}(\beta_0) - \hat{\gamma}^{(i)}(\beta_0)) = o_p(n^{-1/2}). \end{cases} \tag{A.13}$$

For $j = 1, ..., n$, put

$$\begin{aligned} W_{j1} &= \sum_{l=1}^{p}(\hat{\gamma}_l(\beta_0) - \hat{\gamma}_l^{(j)}(\beta_0)) \sum_{t=1}^{n} U_{t-l} \left(\frac{X_{t-1}}{\sqrt{1 + X_{t-1}^2}} + \sum_{k=1}^{p} \gamma_{0,k} \frac{X_{t-k-1}}{\sqrt{1 + X_{t-k-1}^2}} \right) \\ &+ \sum_{l=1}^{p}(\hat{\gamma}_l(\beta_0) - \hat{\gamma}_l^{(j)}(\beta_0)) \sum_{t=1}^{n} \varepsilon_t \frac{X_{t-l-1}}{\sqrt{1 + X_{t-l-1}^2}} \\ &+ \frac{1}{2} \sum_{k=1}^{p} \sum_{l=1}^{p} (\hat{\gamma}_k(\beta_0) - \hat{\gamma}_k^{(j)}(\beta_0))(\hat{\gamma}_l(\beta_0) - \hat{\gamma}_l^{(j)}(\beta_0)) \sum_{t=1}^{n} U_{t-k} \frac{X_{t-l-1}}{\sqrt{1 + X_{t-l-1}^2}} \end{aligned}$$

and

$$\begin{aligned} W_{j2} &= \Bigg(U_{j-1} \{ \frac{X_{j-1}}{\sqrt{1 + X_{j-1}^2}} + \sum_{k=1}^{p} \gamma_{0,k} \frac{X_{j-k-1}}{\sqrt{1 + X_{j-k-1}^2}} \} + \varepsilon_j \frac{X_{j-1-1}}{\sqrt{1 + X_{j-1-1}^2}}, \\ &\quad \cdots, U_{j-p} \{ \frac{X_{j-1}}{\sqrt{1 + X_{j-1}^2}} + \sum_{k=1}^{p} \gamma_{0,k} \frac{X_{j-k-1}}{\sqrt{1 + X_{j-k-1}^2}} \} + \varepsilon_j \frac{X_{j-p-1}}{\sqrt{1 + X_{j-p-1}^2}} \Bigg)^T \end{aligned}$$

Then it follows from Lemma 4, (A.13) and Taylor expansions that

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z^*_{n,i}(\beta_0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{i1} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{\gamma}^{(i)}(\beta_0)-\gamma_0)^T W_{i2} \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i\left(\frac{X_{i-1}}{\sqrt{1+X_{i-1}^2}} + \sum_{k=1}^{p}\gamma_{0,k}\frac{X_{i-k-1}}{\sqrt{1+X_{i-k-1}^2}}\right) + o_p(1) \\
&= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(A^{(i)})^T\Sigma^{-1}W_{i2} \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i\left(\frac{X_{i-1}}{\sqrt{1+X_{i-1}^2}} + \sum_{k=1}^{p}\gamma_{0,k}\frac{X_{i-k-1}}{\sqrt{1+X_{i-k-1}^2}}\right) + o_p(1) \\
&= -(\sqrt{n}A)^T\Sigma^{-1}\frac{1}{n}\sum_{i=1}^{n} W_{i2} \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i\left(\frac{X_{i-1}}{\sqrt{1+X_{i-1}^2}} + \sum_{k=1}^{p}\gamma_{0,k}\frac{X_{i-k-1}}{\sqrt{1+X_{i-k-1}^2}}\right) + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i\tilde{e}_i + o_p(1).
\end{aligned}$$

Now apply a Martingale central limit theorem argument as in Lemma 1 to $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i\tilde{e}_i$ to achieve $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z^*_{n,i}(\beta_0) \xrightarrow{d} N(0, E(\varepsilon_1^2\tilde{e}_1^2))$. Similarly, we can show that

$$\frac{1}{n}\sum_{i=1}^{n} Z^{*2}_{n,i}(\beta_0) = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2\tilde{e}_i^2 + o_p(1) \xrightarrow{p} E(\varepsilon_1^2\tilde{e}_1^2).$$

This completes the proof. □

Proof of Theorem 3. The claim can be proven by using Lemma 5, and arguments in Qin and Lawless (1994). □

Proof of Theorem 4. The argument is similar to the proof of Theorem 3. □

Appendix B

PROOFS OF CHAPTER 4

B.1 Proofs for GARCH(1, 1) Case

In this subsection, we define $\Theta = \{\theta : ||\theta - \theta_0|| \le n^{-\frac{1}{2\gamma}}\}$ for some $1 < \gamma < \min\{\delta_0/2, 2\}$, where $||\cdot||$ denotes the L_2 norm.

Lemma 6. *Under conditions of Theorem 5,*

$$\sup_{1\le t\le n}\sup_{\theta\in\Theta}||\bar{Z}_t(\theta, \alpha_0)|| = o_p(n^{\frac{1}{2\gamma}}). \tag{B.1}$$

Proof. Put $g(t,\theta) = (\bar{Z}_{t,2}(\theta), \bar{Z}_{t,3}(\theta), \bar{Z}_{t,4}(\theta))^T =: (g_1(t,\theta), g_2(t,\theta), g_3(t,\theta))^T$ and

$$\begin{aligned} h(t,\theta) &= \Big(\frac{1-a^t}{1-a}, \frac{\omega\{(1-a^t) - ta^{t-1}(1-a)\}}{(1-a)^2} + b\sum_{k=0}^{t-1} k a^{k-1} Y_{t-k-1}^2, \sum_{k=0}^{t-1} a^k Y_{t-k-1}^2\Big)^T \\ &=: (h_1(t,\theta), h_2(t,\theta), h_3(t,\theta))^T. \end{aligned}$$

Then we have

$$g(t,\theta) = \mathrm{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\} h(t,\theta)\bar{\sigma}_t^{-2}(\theta). \tag{B.2}$$

Write $\sigma_0 = \sigma_0(\theta_0)$. It follows from (5.7) that

$$\begin{aligned} \bar{\sigma}_t^2(\theta) - \sigma_t^2(\theta_0) &= (\omega - \omega_0)\frac{1-a_0^t}{1-a_0} + \omega\left\{\frac{a_0^t - a^t}{1-a} + \frac{(1-a_0^t)(a-a_0)}{(1-a)(1-a_0)}\right\} \\ &\quad + b\sum_{i=0}^{t-1}(a^i - a_0^i)Y_{t-1-i}^2 + (b-b_0)\sum_{i=0}^{t-1} a_0^i Y_{t-1-i}^2 - a_0^t\sigma_0^2. \end{aligned} \tag{B.3}$$

Thus, there exists $C_1 > 0$ such that

$$\left|\bar{\sigma}_t^2(\theta) - \sigma_t^2(\theta_0)\right|/\sigma_t^2(\theta_0) \le [C_1 n^{-\frac{1}{2\gamma}}\{1 + \sum_{i=0}^{t-1} i(\max(a,a_0))^{i-1}Y_{t-1-i}^2\} + a_0^t\sigma_0^2]/\sigma_t^2(\theta_0) \tag{B.4}$$

uniformly for $t \geq 1$ and $\theta \in \Theta$. By the inequality

$$x/(1+x) \leq x^{\tau} \quad \text{for} \quad x > 0 \quad \text{and} \quad 0 < \tau < 1,$$

it can be shown that there exist constants C_2 and $\rho \in (0,1)$ such that for any $\tau \in (0,1)$

$$\sup_{\theta\in\Theta}\{1+\sum_{i=0}^{t-1} i(\max(a,a_0))^{i-1}Y_{t-1-i}^2\}/\sigma_t^2(\theta_0) \leq C_2\Big(1+\sum_{i=0}^{\infty}\rho^i|Y_{t-1-i}|\Big)^{\tau}, \qquad \text{(B.5)}$$

$$a_0^t\sigma_0^2/\sigma_t^2(\theta_0) \leq C_2\sum_{i=t}^{\infty} a_0^{i\tau}|Y_{t-1-i}|^{\tau} \text{ and} \qquad \text{(B.6)}$$

$$\sup_{\theta\in\Theta}||h(t,\theta)||/\bar{\sigma}_t^2(\theta) \leq C_2\Big(1+\sum_{i=0}^{\infty}\rho^i|Y_{t-1-i}|\Big)^{\tau} \qquad \text{(B.7)}$$

uniformly for $t \geq 1$. By (B.4), (B.5) and $a_0 < 1$ (see Remark 3), for any $0 < \delta < 1/2$ there exists a t_δ such that

$$\sup_{t\geq t_\delta}\sup_{\theta\in\Theta}\left|\bar{\sigma}_t^2(\theta)-\sigma_t^2(\theta_0)\right|/\sigma_t^2(\theta_0) < \delta \;\text{ in probability .} \qquad \text{(B.8)}$$

Thus, by inequality $|\log(1+x)| \leq 2|x|$ for all $x > -1/2$, we have for all $t \geq t_\delta$,

$$\begin{aligned}
\left|\log(\bar{\sigma}_t^2(\theta)/\sigma_t^2(\theta_0))\right| &= \left|\log\{1+(\bar{\sigma}_t^2(\theta)-\sigma_t^2(\theta_0))/\sigma_t^2(\theta_0)\}\right| \\
&\leq 2\left|\bar{\sigma}_t^2(\theta)-\sigma_t^2(\theta_0)\right|/\sigma_t^2(\theta_0) \\
&\leq 2C_2\left[C_1 n^{-\frac{1}{2\gamma}}\Big(1+\sum_{i=0}^{\infty}\rho^i|Y_{t-1-i}|\Big)^{\tau}+\Big(\sum_{i=t}^{\infty}a_0^i|Y_{t-1-i}|\Big)^{\tau}\right] \\
&=: d(n,t). \qquad \text{(B.9)}
\end{aligned}$$

It follows that

$$
\begin{aligned}
& \left|\operatorname{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\} - \operatorname{sgn}\{\log \varepsilon_t^2\}\right| \\
= & \left|\operatorname{sgn}\{\log(Y_t^2/\sigma_t^2(\theta_0)) - \log(\bar{\sigma}_t^2(\theta)/\sigma_t^2(\theta_0))\} - \operatorname{sgn}\{\log(Y_t^2/\sigma_t^2(\theta_0))\}\right| \\
\leq & \; 2|I\{\log(\bar{\sigma}_t^2(\theta)/\sigma_t^2(\theta_0)) < \log \varepsilon_t^2 \leq 0\} - I\{0 \leq \log \varepsilon_t^2 < \log(\bar{\sigma}_t^2(\theta)/\sigma_t^2(\theta_0))\}| \\
\leq & \; 2I\{|\log \varepsilon_t^2| \leq d(n,t)\}. \qquad \text{(B.10)}
\end{aligned}
$$

This, combining with (B.7) and (B.9), yields that for any $0 < \tau < 1$,

$$
\begin{aligned}
||g(t,\theta)|| \leq & \; C_2\Big(1 + \sum_{i=0}^{\infty} \rho^i |Y_{t-1-i}|\Big)^{\tau} I\{|\log \varepsilon_t^2| \leq d(n,t)\} \\
& + C_2\Big(1 + \sum_{i=0}^{\infty} \rho^i |Y_{t-1-i}|\Big)^{\tau} \\
=: & \; I_1(t) + I_2(t) \qquad \text{(B.11)}
\end{aligned}
$$

uniformly for $t \geq t_\delta$. By the corollary on p.322 of Nelson (1990), we have $\mathrm{E}\sigma_0^p < \infty$ for any $0 < p < \alpha_0$.

Since $\log \varepsilon_t^2$ is independent of $\mathcal{F}_{t-1} = \sigma(Y_{t-1}, \ldots, Y_{-\infty})$ and its density is continuous at zero, by taking τ small enough such that $\mathrm{E}|Y_t|^{8\gamma\tau} < \infty$ and δ small enough such that $\sup_{|x|\leq\delta} f(x) \leq 2f(0)$, there exists $C_3 > 0$ such that for all $t > t_\delta$,

$$
\begin{aligned}
& \mathrm{E}\left|\Big(1 + \sum_{i=0}^{\infty} \rho^i |Y_{t-1-i}|\Big)^{\tau} I\{|\log \varepsilon_t^2| \leq d(n,t)\}\right|^{4\gamma} \\
= & \; \mathrm{E}\left[\Big(1 + \sum_{i=0}^{\infty} \rho^i |Y_{t-1-i}|\Big)^{\tau} \mathrm{E}\left(I\{|\log \varepsilon_t^2| \leq d(n,t)\}|\mathcal{F}_{t-1}\right)\right]^{4\gamma} \\
\leq & \; C_3 f^{4\gamma}(0) \mathrm{E}\Big(1 + \sum_{i=0}^{\infty} \tilde{\rho}^i |Y_{t-1-i}|\Big)^{8\gamma\tau} \left[n^{-2} + a_0^{4\gamma t\tau}\right], \qquad \text{(B.12)}
\end{aligned}
$$

where $\tilde{\rho} = \max\{\rho, a_0\}$. Therefore, for any $\zeta > 0$,

$$\begin{aligned} & P\{\sup_{1\le t\le n} I_1(t) > \zeta n^{\frac{1}{2\gamma}}\} \\ \le\ & \sum_{t=1}^{t_\delta} P\left\{C_2\Big(1+\sum_{i=0}^{\infty}\rho^i|Y_{t-1-i}|\Big)^{\tau} > \zeta n^{\frac{1}{2\gamma}}\right\} + \sum_{t=t_\delta+1}^{n} \zeta^{-4\gamma} n^{-2}\mathrm{E}|I_1(t)|^{4\gamma} \\ \to\ & 0. \end{aligned} \tag{B.13}$$

Similarly, for any $\zeta > 0$,

$$P\{\sup_{1\le t\le n} I_2(t) > \zeta n^{\frac{1}{2\gamma}}\} \le \sum_{t=1}^{n} \zeta^{-4\gamma} n^{-2}\mathrm{E}|I_2(t)|^{4\gamma} \le C_3\zeta^{-4\gamma}n^{-1} \to 0. \tag{B.14}$$

So it follows from (B.13) and (B.14) that

$$\sup_{1\le t\le n}\sup_{\theta\in\Theta} ||g(t,\theta)|| = o_p(n^{\frac{1}{2\gamma}}). \tag{B.15}$$

On the other hand, similar to (B.5), we have

$$\begin{aligned} & (a + bY_t^2/\bar{\sigma}_t^2(\theta))^{\alpha_0/2} \\ =\ & \{a_0 + b_0Y_t^2/\bar{\sigma}_t^2(\theta) + (a-a_0) + (b-b_0)Y_t^2/\bar{\sigma}_t^2(\theta)\}^{\alpha_0/2} \\ \le\ & \{a_0 + b_0\varepsilon_t^2 + Cn^{-\frac{1}{2\gamma}}(1+\varepsilon_t^2(1+\textstyle\sum_{i=0}^{\infty}\rho^i|Y_{t-1-i}|)^{\tau} + \omega^{-1}a_0^t\varepsilon_t^2\sigma_0^2)\}^{\alpha_0/2} \end{aligned} \tag{B.16}$$

uniformly for $t \ge 1$. Thus, by the inequality: $(1+x)^p \le 1 + 2px$ for $p > 0$ and small $x > 0$, we have

$$\sup_{1\le t\le n}\sup_{\theta\in\Theta} |\{a + bY_t^2/\bar{\sigma}_t^2(\theta)\}^{\alpha_0/2} - (a_0 + b_0\varepsilon_t^2)^{\alpha_0/2}| = o_p(n^{\frac{1}{2\gamma}}). \tag{B.17}$$

Since $\{\varepsilon_t^2\}$ is a sequence of independent and identically distributed random variables with $\mathrm{E}|\varepsilon_t|^{\delta_0} < \infty$ and $\mathrm{E}(a_0 + b_0\varepsilon_t^2)^{\alpha_0/2} = 1$, we have

$$\sup_{1\le t\le n} |(a_0 + b_0\varepsilon_t^2)^{\alpha_0/2} - 1| = o_p(n^{\frac{1}{2\gamma}}). \tag{B.18}$$

Thus, the lemma follows from (B.15), (B.17) and (B.18). □

Lemma 7. *Under Conditions of Theorem 5,*

$$(i) \qquad \sup_{\theta\in\Theta} \Big|\Big|\frac{1}{n}\sum_{t=1}^{n} \bar{Z}_t(\theta,\alpha_0)\bar{Z}_t^T(\theta,\alpha_0) - \mathrm{E}\{Z_1(\theta_0,\alpha_0)Z_1^T(\theta_0,\alpha_0)\}\Big|\Big| = o_p(1), \text{ and}$$

$$(ii) \qquad \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \bar{Z}_t(\theta_0,\alpha_0) \xrightarrow{d} \mathrm{N}\Big(0, \mathrm{E}\{Z_1(\theta_0,\alpha_0)Z_1^T(\theta_0,\alpha_0)\}\Big),$$

where $\bar{Z}_{t,i}(\theta,\alpha_0) = \bar{Z}_{t,i}(\theta)$ *when* $i = 2,3,4$, *and* $Z_t(\theta,\alpha_0)$ *is defined as* $\bar{Z}_t(\theta,\alpha_0)$ *with* $\sigma_t^2(\theta)$ *replaced by* $\bar{\sigma}_t^2(\theta)$.

Proof. For the proof of (i), it is sufficient to show that

$$\sup_{\theta\in\Theta} \Big|\Big|\frac{1}{n}\sum_{t=1}^{n}\{\bar{Z}_{t,i}(\theta,\alpha_0)\bar{Z}_{t,j}(\theta,\alpha_0)\} - \mathrm{E}\{Z_{t,i}(\theta_0,\alpha_0)Z_{t,j}(\theta_0,\alpha_0)\}\Big|\Big| = o_p(1) \qquad \text{(B.19)}$$

for $i, j = 1,2,3,4$. Here we only show the case of $i = 3$ and $j = 4$, since the other cases can be proved similarly. Define $\widetilde{h}_2(t,\theta) = \omega/(1-a)^2 + b\sum_{k=0}^{\infty} ka^{k-1}Y_{t-k-1}^2$. By (B.2),

$$\begin{aligned}
\bar{Z}_{t,3}(\theta,\alpha_0) \quad = \quad & \mathrm{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\}\Big(\frac{\omega}{(1-a)^2} + b\sum_{k=0}^{\infty} ka^{k-1}Y_{t-k-1}^2\Big)\bar{\sigma}_t^{-2}(\theta) \\
& -\mathrm{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\}\Big(\frac{\omega a^t + ta^{t-1}(1-a)}{(1-a)^2} + b\sum_{k=t}^{\infty} ka^{k-1}Y_{t-k-1}^2\Big)\bar{\sigma}_t^{-2}(\theta) \\
= \quad & \mathrm{sgn}(\log\varepsilon_t^2)\,\Big[\widetilde{h}_2(t,\theta_0)/\sigma_t^2(\theta_0)\Big] \\
& +\mathrm{sgn}(\log\varepsilon_t^2)\,\Big[(\widetilde{h}_2(t,\theta) - \widetilde{h}_2(t,\theta_0))/\sigma_t^2(\theta_0)\Big] \\
& +\mathrm{sgn}(\log\varepsilon_t^2)\,\big[(\sigma_t^2(\theta_0) - \bar{\sigma}_t^2(\theta))/\sigma_t^2(\theta_0)\big]\,\Big[\widetilde{h}_2(t,\theta)/\bar{\sigma}_t^2(\theta)\Big] \\
& -\mathrm{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\}a^{t-1}\left[\Big(\frac{\omega a + t(1-a)}{(1-a)^2} + b\sum_{k=0}^{\infty}(t+k)a^kY_{-k-1}^2\Big)/\bar{\sigma}_t^2(\theta)\right] \\
& + \Big[\mathrm{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\} - \mathrm{sgn}(\log\varepsilon_t^2)\Big]\,\Big[\widetilde{h}_2(t,\theta)/\bar{\sigma}_t^2(\theta)\Big] \\
=: \quad & L_1(t) + L_2(t) + L_3(t) + L_4(t) + L_5(t)
\end{aligned}$$

and

$$\begin{aligned}
\bar{Z}_{t,4}(\theta,\alpha_0) &= \operatorname{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\}\sum_{k=0}^{t-1}a^kY_{t-k-1}^2/\bar{\sigma}_t^2(\theta)\\
&= \operatorname{sgn}(\log\varepsilon_t^2)\sum_{k=0}^{\infty}a_0^kY_{t-k-1}^2/\sigma_t^2(\theta_0)\\
&\quad+\operatorname{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\}\sum_{k=0}^{\infty}[a^kY_{t-k-1}^2/\sigma_t^2(\theta_0)][(\sigma_t^2(\theta_0)-\bar{\sigma}_t^2(\theta))/\bar{\sigma}_t^2(\theta)]\\
&\quad+\operatorname{sgn}(\log\varepsilon_t^2)\sum_{k=0}^{\infty}(a^k-a_0^k)Y_{t-k-1}^2/\sigma_t^2(\theta_0)\\
&\quad-\operatorname{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\}\sum_{k=t}^{\infty}a^kY_{t-k-1}^2/\bar{\sigma}_t^2(\theta)\\
&\quad+\left[\operatorname{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta))\}-\operatorname{sgn}(\log\varepsilon_t^2)\right]\sum_{k=0}^{\infty}a^kY_{t-k-1}^2/\sigma_t^2(\theta_0)\\
&=: M_1(t)+M_2(t)+M_3(t)+M_4(t)+M_5(t).
\end{aligned}$$

Similar to (B.5) and (B.7), there exist $C_4>0$ and $\rho\in(0,1)$ such that for any $0<\tau<1$,

$$\begin{aligned}
&\sup_{\theta\in\Theta}|L_2(t)+L_3(t)+L_4(t)|\\
\le\;& \sup_{\theta\in\Theta}C_4\Big\{\Big[n^{-\frac{1}{2\gamma}}\Big(1+\sum_{i=0}^{\infty}\rho^i|Y_{t-i-1}|\Big)^{\tau}+a_0^t\sigma_0^2\Big]\Big(1+\sum_{i=0}^{\infty}\rho^i|Y_{t-i-1}|\Big)^{\tau}\Big\}\\
&+\omega^{-1}a^{t-1}\left[\Big(\frac{\omega a+t(1-a)}{(1-a)^2}+b\sum_{k=0}^{\infty}(t+k)a^kY_{-k-1}^2\Big)\right]\\
\le\;& C_4\Big[n^{-\frac{1}{2\gamma}}\Big(1+\sum_{i=0}^{\infty}\rho^i|Y_{t-i-1}|\Big)^{\tau}+a_0^t\sigma_0^2\Big]\Big(1+\sum_{i=0}^{\infty}\rho^i|Y_{t-i-i}|\Big)^{\tau} \qquad\text{(B.20)}\\
&+2\omega_0^{-1}\left(\frac{a_0+1}{2}\right)^{t-1}\left\{\frac{4\omega_0(1+a_0)+6t(1-a_0)}{(1-a_0)^2}+2b_0\sum_{k=0}^{\infty}(t+k)\left(\frac{a_0+1}{2}\right)^kY_{-k-1}^2\right\},
\end{aligned}$$

where the inequalities follow by taking n sufficiently large such that $C_4n^{-\frac{1}{2\gamma}}\le\min\{(1-$

$a_0)/2, \omega_0/2, b_0/2\}$, and

$$\begin{aligned}
&\sup_{\theta\in\Theta} |M_2(t) + M_3(t) + M_4(t)| \\
\leq\ & C_4\Big\{n^{-\frac{1}{2\gamma}}\Big(1+\sum_{i=0}^{\infty}\rho^i|Y_{t-i-1}|\Big)^{\tau} + a_0^t\sigma_0^2\Big\}\Big(1+\sum_{i=0}^{\infty}\rho^i|Y_{t-i-1}|\Big)^{\tau} \\
&+2\omega_0^{-1}\left(\frac{a_0+1}{2\gamma}\right)^t\sum_{k=0}^{\infty}\left(\frac{a_0+1}{2}\right)^k Y_{-k-1}^2. \qquad \text{(B.21)}
\end{aligned}$$

Thus, by (B.20) and (B.21), we can show that

$$\frac{1}{n}\sum_{t=1}^{n}\left\{\sup_{\theta\in\Theta}[|M_1(t)|(|L_2(t)+L_3(t)+L_4(t)|)]\right\}\xrightarrow{p}0, \qquad \text{(B.22)}$$

$$\frac{1}{n}\sum_{t=1}^{n}\left\{\sup_{\theta\in\Theta}[|L_1(t)|(|M_2(t)+M_3(t)+M_4(t)|)]\right\}\xrightarrow{p}0,\ \text{and} \qquad \text{(B.23)}$$

$$\frac{1}{n}\sum_{t=1}^{n}\left\{\sup_{\theta\in\Theta}[(|L_2(t)+L_3(t)+L_4(t)|)(|M_2(t)+M_3(t)+M_4(t)|)]\right\}\xrightarrow{p}0. \qquad \text{(B.24)}$$

Further, using the same arguments as in proving Lemma 5.1, we have

$$\frac{1}{n}\sum_{t=1}^{n}\left\{\sup_{\theta\in\Theta}[|M_5(t)|(|L_1(t)+L_2(t)+L_3(t)+L_4(t)+L_5(t)|)]\right\}\xrightarrow{p}0,\ \text{and} \qquad \text{(B.25)}$$

$$\frac{1}{n}\sum_{t=1}^{n}\left\{\sup_{\theta\in\Theta}[|L_5(t)|(|M_1(t)+M_2(t)+M_3(t)+M_4(t)|)]\right\}\xrightarrow{p}0. \qquad \text{(B.26)}$$

Therefore,

$$\begin{aligned}
&\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}\{\bar{Z}_{t,3}(\theta,\alpha_0)\bar{Z}_{t,4}(\theta,\alpha_0)\} - \mathrm{E}\{Z_{t,3}(\theta_0,\alpha_0)Z_{t,4}(\theta_0,\alpha_0)\}\right| \\
=\ &\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}\{L_1(t)M_1(t)\} - \mathrm{E}\{Z_{t,3}(\theta_0,\alpha_0)Z_{t,4}(\theta_0,\alpha_0)\}\right| + o_p(1) = o_p(1), \qquad \text{(B.27)}
\end{aligned}$$

i.e., (B.19) holds for the case of $i = 3$ and $j = 4$.

Next, we prove (ii). Note that

$$\begin{aligned} g(t,\theta_0) &= \operatorname{sgn}\{\log \varepsilon_t^2\} h(t,\theta_0)/\sigma_t^2(\theta_0) \\ &\quad + \left(\operatorname{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta_0))\} - \operatorname{sgn}\{\log \varepsilon_t^2\}\right) h(t,\theta_0)/\sigma_t^2(\theta_0) \\ &\quad + \operatorname{sgn}\{\log(Y_t^2/\bar{\sigma}_t^2(\theta_0))\} h(t,\theta_0)[1/\bar{\sigma}_t^2(\theta_0) - 1/\sigma_t^2(\theta_0)] \\ &=: H_1(t) + H_2(t) + H_3(t). \end{aligned} \tag{B.28}$$

Using $\sigma_t^2(\theta_0) - \bar{\sigma}_t^2(\theta_0) = a_0^t \sigma_0^2$ and the same arguments as in deriving (B.10)–(B.12), we have

$$\mathrm{E}\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(||H_2(t)|| + ||H_3(t)||)\}^p = O\left\{n^{-p/2}\sum_{t=1}^{n}\mathrm{E}||a_0^t\sigma_0^2 h(t,\theta_0)||^p\right\} = O(n^{-p/2}) \tag{B.29}$$

for any $0 < p < \min\{1, \alpha_0/8\}$, which implies that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} g(t,\theta_0) - \frac{1}{\sqrt{n}}\sum_{t=1}^{n} H_1(t) \xrightarrow{p} 0. \tag{B.30}$$

Similarly, we can show that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{(a_0 + b_0 Y_t^2/\bar{\sigma}_t^2(\theta_0))^{\alpha_0/2} - 1\} - \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{(a_0 + b_0\varepsilon_t^2)^{\alpha_0/2} - 1\} \xrightarrow{p} 0. \tag{B.31}$$

Note that $H_1(t) = (Z_{t,2}(\theta_0), Z_{t,3}(\theta_0), Z_{t,4}(\theta_0))^T$. By (B.30) and (B.31), we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\bar{Z}_t(\theta_0,\alpha_0) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} Z_t(\theta_0,\alpha_0) + o_p(1). \tag{B.32}$$

Since $Z_t(\theta_0, \alpha_0)$ is a martingale difference sequence, (ii) follows from (B.32) and the central limit theorem (CLT) for martingales (see Hall and Heyde (1980)). This completes the proof of Lemma 5.2. □

Lemma 8. *Under conditions of Theorem 5,*

$$\sup_{\theta\in\Theta}\Big|\Big|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big\{\Big[sgn\Big(\log\frac{Y_t^2}{\bar{\sigma}_t^2(\theta)}\Big)-sgn\Big(\log\frac{Y_t^2}{\bar{\sigma}_t^2(\theta_0)}\Big)\Big]\frac{h(t,\theta)}{\bar{\sigma}_t^2(\theta)}$$
$$-2f(0)\Big(\frac{\sigma_t^2(\theta_0)-\bar{\sigma}_t^2(\theta)}{\sigma_t^2(\theta_0)}\Big)\frac{h(t,\theta_0)}{\bar{\sigma}_t^2(\theta_0)}\Big\}\Big|\Big|=o_p(1),$$

where $h(t,\theta)$ is defined in the proof of Lemma 5.1.

Proof. By Taylor expansion, similar to Lemma 5.2, it can be shown that

$$\sup_{\theta\in\Theta}\Big|\Big|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big\{\Big[\mathrm{sgn}\Big(\log\frac{Y_t^2}{\bar{\sigma}_t^2(\theta)}\Big)-\mathrm{sgn}\Big(\log\frac{Y_t^2}{\bar{\sigma}_t^2(\theta_0)}\Big)\Big]\Big(\frac{h(t,\theta)}{\bar{\sigma}_t^2(\theta)}-\frac{h(t,\theta_0)}{\bar{\sigma}_t^2(\theta_0)}\Big)\Big\}\Big|\Big|=o_p(1).$$

Further, similar to (B.30), we have

$$\sup_{\theta\in\Theta}\Big|\Big|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[\mathrm{sgn}\left(\log\frac{Y_t^2}{\bar{\sigma}_t^2(\theta_0)}\right)-\mathrm{sgn}(\log\varepsilon_t^2)\right]\frac{h(t,\theta_0)}{\bar{\sigma}_t^2(\theta_0)}\Big|\Big|=o_p(1).$$

Thus, for proving Lemma 5.3, it suffices to show that

$$\sup_{\theta\in\Theta}\Big|\Big|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big[\mathrm{sgn}\Big(\log\frac{Y_t^2}{\bar{\sigma}_t^2(\theta)}\Big)-\mathrm{sgn}(\log\varepsilon_t^2)-2f(0)\Big(\frac{\sigma_t^2(\theta_0)-\bar{\sigma}_t^2(\theta)}{\sigma_t^2(\theta_0)}\Big)\Big]\frac{h(t,\theta_0)}{\bar{\sigma}_t^2(\theta_0)}\Big|\Big|=o_p(1).$$

Put $\xi_{1t}(\theta)=[2I(\log(\frac{\bar{\sigma}_t^2(\theta)}{\sigma_t^2(\theta_0)})<\log\varepsilon_t^2<0)+I(\log(\frac{\bar{\sigma}_t^2(\theta)}{\sigma_t^2(\theta_0)})<\log\varepsilon_t^2=0)]h(t,\theta_0)/\bar{\sigma}_t^2(\theta_0)$, $\xi_{2t}(\theta)=[2I(0<\log\varepsilon_t^2\leq\log(\frac{\bar{\sigma}_t^2(\theta)}{\sigma_t^2(\theta_0)}))+I(0=\log\varepsilon_t^2\leq\log(\frac{\bar{\sigma}_t^2(\theta)}{\sigma_t^2(\theta_0)}))]h(t,\theta_0)/\bar{\sigma}_t^2(\theta_0)$ and $\mathcal{F}_t=\sigma(\varepsilon_s,\ s\leq t)$. Then $\Big[\mathrm{sgn}\Big(\log\frac{Y_t^2}{\bar{\sigma}_t^2(\theta)}\Big)-\mathrm{sgn}(\log\varepsilon_t^2)\Big]\frac{h^T(t,\theta_0)}{\bar{\sigma}_t^2(\theta_0)}=\xi_{1t}(\theta)-\xi_{2t}(\theta)$ and

$$\begin{aligned}\mathrm{E}[(\xi_{1t}(\theta)-\xi_{2t}(\theta))|\mathcal{F}_{t-1}] &= -2f(0)\log\left(\bar{\sigma}_t^2(\theta)/\sigma_t^2(\theta_0)\right)\left(h(t,\theta_0)/\bar{\sigma}_t^2(\theta_0)\right)(1+o_p(1))\\ &= 2f(0)\left(\frac{\sigma_t^2(\theta_0)-\bar{\sigma}_t^2(\theta)}{\sigma_t^2(\theta_0)}\right)\frac{h(t,\theta_0)}{\bar{\sigma}_t^2(\theta_0)}(1+o_p(1))\end{aligned}$$

holds uniformly in $\theta\in\Theta$. Hence, we only need to show that

$$\sup_{\theta\in\Theta}\Big\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{\xi_{it}(\theta)-\mathrm{E}[\xi_{it}(\theta)|\mathcal{F}_{t-1}]\}\Big\|=o_p(1)\text{ for }i=1,2. \qquad \text{(B.33)}$$

It follows from (B.7) and (B.10) that for $i = 1, 2$

$$\frac{1}{n}\sum_{t=1}^{n} \mathrm{E}\|\xi_{it}(\theta)\|^2 = o(1) \text{ and } \frac{1}{n}\sum_{t=1}^{n} \mathrm{E}\|\xi_{it}(\theta_1) - \xi_{it}(\theta_2)\|^2 \le C||\theta_1 - \theta_2||. \tag{B.34}$$

Note that for any given θ, $\{\xi_{it}(\theta) - \mathrm{E}[\xi_{it}(\theta)|\mathcal{F}_{t-1}]\}$ is a martingale difference sequence. By (B.34) and a chaining technique (see pp. 356-358 in Hansen (1996) or pp. 330–331 in Koul and Surgailis (2001)), (B.33) can be derived. □

Proof of Theorem 2.1. By Theorem 1 of Peng and Yao (2003), under Conditions 1 and 2, there exists a positive definition matrix Ω such that

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Omega). \tag{B.35}$$

Thus, by (B.16) with α instead of α_0 and some similar arguments as in proving (B.17), we have for any $0 \le \alpha \le \delta_0$,

$$\frac{1}{n}\sum_{t=1}^{n}\{(\hat{a} + \hat{b}\bar{\varepsilon}_t^2(\hat{\theta}))^{\alpha/2} - (a_0 + b_0\varepsilon_t^2)^{\alpha/2}\} \xrightarrow{p} 0. \tag{B.36}$$

It follows from the weak law of large numbers that

$$\frac{1}{n}\sum_{t=1}^{n}\{(a_0 + b_0\varepsilon_t^2)^{\alpha/2} - \mathrm{E}(a_0 + b_0\varepsilon_t^2)^{\alpha/2}\} \xrightarrow{p} 0.$$

Since the convergence of a monotone function to its limit is uniform over any closed interval, we have

$$\sup_{0\le\alpha\le\delta_0}\left|\frac{1}{n}\sum_{t=1}^{n}\{(\hat{a} + \hat{b}\bar{\varepsilon}_t^2(\hat{\theta}))^{\alpha/2} - \mathrm{E}(a_0 + b_0\varepsilon_t^2)^{\alpha/2}\}\right| \xrightarrow{p} 0. \tag{B.37}$$

Since α_0 is the unique positive solution to $\mathrm{E}(a_0 + b_0\varepsilon_t^2)^{\alpha/2} = 1$, we have $\hat{\alpha} \xrightarrow{p} \alpha_0$. Thus, by

Taylor expansion, when $|\hat{\alpha}-\alpha_0|\le \nu$ with $\nu>0$ small enough,

$$\begin{aligned}
0 &= \frac{1}{n}\sum_{t=1}^{n}(\hat{a}+\hat{b}\bar{\varepsilon}_t^2(\hat{\theta}))^{\hat{\alpha}/2}-1\\
&= \frac{1}{n}\sum_{t=1}^{n}[(\hat{a}+\hat{b}\bar{\varepsilon}_t^2(\hat{\theta}))^{\hat{\alpha}/2}-(a_0+b_0\varepsilon_t^2)^{\alpha_0/2}]+\frac{1}{n}\sum_{t=1}^{n}[(a_0+b_0\varepsilon_t^2)^{\alpha_0/2}-1]\\
&= \frac{\hat{\alpha}}{2n}\sum_{t=1}^{n}[\tilde{a}_0+\tilde{b}_0\tilde{\varepsilon}_t^2]^{\hat{\alpha}/2-1}[(\hat{a}-a_0)+(\hat{b}\bar{\varepsilon}_t^2(\hat{\theta})-b_0\varepsilon_t^2)]\\
&\quad +\frac{1}{n}\sum_{t=1}^{n}[(a_0+b_0\varepsilon_t^2)^{\alpha_0/2}-1]+\frac{1}{2n}\sum_{t=1}^{n}(a_0+b_0\varepsilon_t^2)^{\alpha_0/2}\log(a_0+b_0\varepsilon_t^2)(\hat{\alpha}-\alpha_0)\\
&\quad +\frac{1}{4n}\sum_{t=1}^{n}(a_0+b_0\varepsilon_t^2)^{\tilde{\alpha}/2}\{\log(a_0+b_0\varepsilon_t^2)\}^2(\hat{\alpha}-\alpha_0)^2, \qquad \text{(B.38)}
\end{aligned}$$

where $(\tilde{a},\tilde{b},\tilde{\varepsilon}_t^2,\tilde{\alpha})$ lies between $(\hat{a},\hat{b},\bar{\varepsilon}_t^2(\hat{\theta}),\hat{\alpha})$ and $(a_0,b_0,\varepsilon_t^2,\alpha_0)$. Thus, like the proof of Lemma 5.2, the right hand side of (B.38) is equal to

$$\begin{aligned}
&\Big\{\frac{\alpha_0}{2n}\sum_{t=1}^{n}[a_0+b_0\varepsilon_t^2]^{\alpha_0/2-1}[(\hat{a}-a_0)+(\hat{b}-b_0)\varepsilon_t^2-(\hat{\theta}-\theta_0)^T(b_0\varepsilon_t^2h(t,\theta_0)/\bar{\sigma}_t^2(\theta_0))]\\
&+\frac{1}{n}\sum_{t=1}^{n}[(a_0+b_0\varepsilon_t^2)^{\frac{\alpha_0}{2}}-1]+\frac{1}{2n}\sum_{t=1}^{n}(a_0+b_0\varepsilon_t^2)^{\frac{\alpha_0}{2}}\log(a_0+b_0\varepsilon_t^2)(\hat{\alpha}-\alpha_0)\Big\}(1+o_p(1))\\
=\;&\frac{\alpha_0}{2}(\hat{a}-a_0)\mathrm{E}[a_0+b_0\varepsilon_t^2]^{\frac{\alpha_0}{2}-1}+\frac{e_0}{2}(\hat{b}-b_0)-\frac{b_0e_0}{2}(\hat{\theta}-\theta_0)^T\lim_{t\to\infty}\mathrm{E}(h(t,\theta_0)/\bar{\sigma}_t^2(\theta_0))\\
&+\frac{1}{n}\sum_{t=1}^{n}[(a_0+b_0\varepsilon_t^2)^{\frac{\alpha_0}{2}}-1]+\frac{1}{2}(\hat{\alpha}-\alpha_0)\mathrm{E}[(a_0+b_0\varepsilon_t^2)^{\frac{\alpha_0}{2}}\log(a_0+b_0\varepsilon_t^2)]+o_p(n^{-\frac{1}{2}}),
\end{aligned}$$

where $e_0=\alpha_0\mathrm{E}[(a_0+b_0\varepsilon_t^2)^{\alpha_0/2-1}\varepsilon_t^2]$. As a result, let $A_0=\mathrm{E}[(a_0+b_0\varepsilon_t^2)^{\frac{\alpha_0}{2}}\log(a_0+b_0\varepsilon_t^2)]$, we have

$$\begin{aligned}
&\sqrt{n}(\hat{\alpha}-\alpha_0)\\
=\;&A_0^{-1}\sqrt{n}\left\{\alpha_0(\hat{a}-\hat{a}_0)\mathrm{E}[a_0+b_0\varepsilon_t^2]^{\frac{\alpha_0}{2}-1}+e_0(\hat{b}-\hat{b}_0)-\frac{b_0e_0}{2}(\hat{\theta}-\theta_0)^T\lim_{t\to\infty}\mathrm{E}\left(\frac{h(t,\theta_0)}{\bar{\sigma}_t^2(\theta_0)}\right)\right\}\\
&+\frac{2A_0^{-1}}{\sqrt{n}}\sum_{t=1}^{n}[(a_0+b_0\varepsilon_t^2)^{\frac{\alpha_0}{2}}-1]+o_p(1). \qquad \text{(B.39)}
\end{aligned}$$

Thus, by (B.35) and the CLT for martingales (see Hall and Heyde (1980)), the right-hand side of (B.39) converges in distribution to a Gaussian distribution with asymptotic variance Σ depending on the asymptotic covariance of $\sqrt{n}(\hat{\theta}-\theta_0)$ and $\frac{1}{\sqrt{n}}\sum_{t=1}^n[(a_0+b_0\varepsilon_t^2)^{\alpha_0/2}-1]$. □

Proof of Theorem 2.2. Put

$$\theta=\theta_0+n^{-1/2}\nu,\quad \nu=(\nu_1,\nu_2,\nu_3)^T \quad\text{and}\quad S_{11}=\mathrm{E}\{Z_1(\theta_0,\alpha_0)Z_1^T(\theta_0,\alpha_0)\}.$$

Then by Lemmas 5.1, 5.2 and some similar arguments as in the proof of Theorem 1 of Owen (1990), we have

$$\begin{aligned}
&l(\theta,\alpha_0)\\
=&\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n \bar{Z}_t(\theta,\alpha_0)\right)^T\left(\frac{1}{n}\sum_{t=1}^n \bar{Z}_t(\theta,\alpha_0)\bar{Z}_t^T(\theta,\alpha_0)\right)\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n \bar{Z}_t(\theta,\alpha_0)\right)(1+o_p(1))\\
=&\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n \bar{Z}_t(\theta,\alpha_0)\right)^T S_{11}^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n \bar{Z}_t(\theta,\alpha_0)\right)(1+o_p(1)), \qquad \text{(B.40)}
\end{aligned}$$

holds uniformly for all $\theta\in\Theta$. Especially,

$$l(\theta_0,\alpha_0) = \left(\frac{1}{\sqrt{n}}\sum_{t=1}^n \bar{Z}_t(\theta_0,\alpha_0)\right)^T S_{11}^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n \bar{Z}_t(\theta_0,\alpha_0)\right)(1+o_p(1)). \qquad \text{(B.41)}$$

Put $\Delta_n(\theta)=\frac{1}{\sqrt{n}}\sum_{t=1}^n\bar{Z}_t(\theta,\alpha_0)$. Then

$$\begin{aligned}
l(\theta,\alpha_0)-l(\theta_0,\alpha_0) &= (\Delta_n(\theta)-\Delta_n(\theta_0))^T S_{11}^{-1}\Delta_n(\theta_0)+\Delta_n^T(\theta_0)S_{11}^{-1}(\Delta_n(\theta)-\Delta_n(\theta_0))\\
&\quad +(\Delta_n(\theta)-\Delta_n(\theta_0))^T S_{11}^{-1}(\Delta_n(\theta)-\Delta_n(\theta_0))+o_p(1) \quad \text{and}\\
\Delta_n(\theta)-\Delta_n(\theta_0) &= \frac{1}{\sqrt{n}}\sum_{t=1}^n\Big\{[(a+bY_t^2/\bar{\sigma}_t^2(\theta))^{\alpha_0/2}-(a_0+b_0Y_t^2/\bar{\sigma}_t^2(\theta_0))^{\alpha_0/2}],\\
&\quad \{\mathrm{sgn}[\log(Y_t^2/\bar{\sigma}_t^2(\theta))]-\mathrm{sgn}[\log(Y_t^2/\bar{\sigma}_t^2(\theta_0))]\}h^T(t,\theta)/\bar{\sigma}_t^2(\theta)\\
&\quad +\mathrm{sgn}[\log(Y_t^2/\bar{\sigma}_t^2(\theta_0))][h^T(t,\theta)/\bar{\sigma}_t^2(\theta)-h^T(t,\theta_0)/\bar{\sigma}_t^2(\theta_0)]\Big\}^T\\
&=: \frac{1}{\sqrt{n}}\sum_{t=1}^n(Z_{t1}(\nu),g_A(t,\nu)+g_B(t,\nu))^T.
\end{aligned}$$

By Taylor expansion (see also the expression for (B.38)), it follows that

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} Z_{t1}(\nu) &= \frac{\alpha_0}{2n}\sum_{t=1}^{n}[a_0+b_0\varepsilon_t^2]^{\frac{\alpha_0}{2}-1}\left[\nu_2+\nu_3\varepsilon_t^2-\nu^T b_0\varepsilon_t^2\frac{h(t,\theta_0)}{\bar{\sigma}_t^2(\theta_0)}\right]+o_p(1)\\ &= \frac{\nu^T\alpha_0}{2n}\sum_{t=1}^{n}[a_0+b_0\varepsilon_t^2]^{\frac{\alpha_0}{2}-1}\Big[-\frac{b_0\varepsilon_t^2 h_1(t,\theta_0)}{\bar{\sigma}_t^2(\theta_0)},\,1-\frac{b_0\varepsilon_t^2 h_2(t,\theta_0)}{\bar{\sigma}_t^2(\theta_0)},\\ &\qquad \varepsilon_t^2-\frac{b_0\varepsilon_t^2 h_3(t,\theta_0)}{\bar{\sigma}_t^2(\theta_0)}\Big]^T(1+o_p(1)) \qquad \text{(B.42)}\end{aligned}$$

holds uniformly for all $\theta\in\Theta$. By Lemma 5.3 and Taylor expansion, we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} g_A(t,\nu) = \frac{-2\nu^T f(0)}{n}\sum_{t=1}^{n}\left(\frac{h(t,\theta_0)}{\sigma_t^2(\theta_0)}\right)\left(\frac{h(t,\theta_0)}{\sigma_t^2(\theta_0)}\right)^T(1+o_p(1)) \text{ and} \qquad \text{(B.43)}$$

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} g_B(t,\nu) &= \frac{\nu^T}{n}\sum_{t=1}^{n}\text{sgn}(\log\varepsilon_t^2)\Big[\sigma_t^{-2}(\theta_0)\Big(\frac{\partial h_1(t,\theta_0)}{\partial\theta},\frac{\partial h_2(t,\theta_0)}{\partial\theta},\frac{\partial h_3(t,\theta_0)}{\partial\theta}\Big)\\ &\quad -\Big(h(t,\theta_0)/\sigma_t^2(\theta_0)\Big)\Big(h(t,\theta_0)/\sigma_t^2(\theta_0)\Big)^T\Big](1+o_p(1)) \qquad \text{(B.44)}\end{aligned}$$

holds uniformly for all $\theta\in\Theta$, where $\frac{\partial h_i(t,\theta_0)}{\partial\theta}=(\frac{\partial h_i(t,\theta_0)}{\partial\theta_1},\frac{\partial h_i(t,\theta_0)}{\partial\theta_2},\frac{\partial h_i(t,\theta_0)}{\partial\theta_3})^T$ for $i=1,2,3$. Since the median of $\text{sgn}(\log\varepsilon_t^2)$ is zero, it follows from the weak law of large numbers for a martingale that the right-hand side of (B.44) converges to zero in probability. Put $d_0 = b_0\text{E}[(a_0+b_0\varepsilon_t^2)^{\alpha_0/2-1}\varepsilon_t^2]$ and define

$$\begin{aligned}A_1 &= \lim_{t\to\infty}\frac{\alpha_0 d_0}{2}\left[-\text{E}\left(\frac{h_1(t,\theta_0)}{\bar{\sigma}_t^2(\theta_0)}\right),\frac{\text{E}[a_0+b_0\varepsilon_t^2]^{\frac{\alpha_0}{2}-1}}{d_0}-\text{E}\left(\frac{h_2(t,\theta_0)}{\bar{\sigma}_t^2(\theta_0)}\right),\text{E}\left(\frac{1}{b_0}-\frac{h_3(t,\theta_0)}{\bar{\sigma}_t^2(\theta_0)}\right)\right]^T,\\ A_2 &= \lim_{t\to\infty}\text{E}\left[-2f(0)\left(h(t,\theta_0)/\sigma_t^2(\theta_0)\right)\left(h(t,\theta_0)/\sigma_t^2(\theta_0)\right)^T\right]\end{aligned}$$

and $A=(A_1,A_2)$. It follows from (B.42)–(B.44) that

$$l(\theta,\alpha_0)-l(\theta_0,\alpha_0)=\left(\nu^T A S_{11}^{-1}\Delta_n(\theta_0)+\Delta_n^T(\theta_0)S_{11}^{-1}A^T\nu+\nu^T A S_{11}^{-1}A^T\nu\right)(1+o_p(1)) \qquad \text{(B.45)}$$

holds uniformly for all $\theta\in\Theta$. Like the proof of Lemma 1 of Qin and Jin (1994), we know that the minimizer $\hat{\theta}=\theta_0+n^{-1/2}\nu$ of (B.45) must lie in Θ. Thus, by minimizing (B.45) with respect to ν, it follows that

$$\hat{\nu} = -(AS_{11}^{-1}A^T)^{-1}AS_{11}^{-1}\Delta_n(\theta_0) + o_p(1).$$

Substitute this into (B.45), we have

$$\begin{aligned} & l(\hat{\theta}, \alpha_0) \\ = \; & [S_{11}^{-1/2}\Delta_n(\theta_0)]^T[I - S_{11}^{-1/2}A^T(AS_{11}^{-1}A^T)^{-1}AS_{11}^{-1/2}][S_{11}^{-1/2}\Delta_n(\theta_0)](1 + o_p(1)). \end{aligned} \tag{B.46}$$

By Lemma 5.2, $S_{11}^{-1/2}\Delta_n(\theta_0)$ converges in distribution to a multivariate standard normal distribution. Thus, by (B.46) and noting that the trace of $I - S_{11}^{-1/2}A^T(AS_{11}^{-1}A^T)^{-1}AS_{11}^{-1/2}$ is 1, we have $l(\hat{\theta}, \alpha_0) \xrightarrow{d} \chi^2(1)$, i.e., Theorem 6 follows. □

B.2 Proofs for AR(1)-ARCH(1) Case

In this subsection, we define $\Theta = \{\theta : ||\theta - \theta_0|| \leq n^{-\frac{1}{2\gamma}}\}$ for some $\gamma \in (1, \alpha_0)$.

Lemma 9. *Under conditions of Theorem 7,*

$$\sup_{1\leq t\leq n} \sup_{\theta\in\Theta} ||X_t(\theta, \alpha_0)|| = o_p(n^{\frac{1}{2\gamma}}). \tag{B.47}$$

Proof. Define $G(t, \theta) = (X_{t2}(\theta), X_{t3}(\theta), X_{t4}(\theta))^T$ and

$$H(t, \theta) = (1, 2(Y_t - aY_{t-1})Y_{t-1}, Y_{t-1}^2)^T/(1 + Y_{t-1}^2).$$

Then $G(t, \theta) = -\text{sgn}(\varepsilon_t^2(\omega, a, b))H(t, \theta)$ and

$$\sup_{\theta\in\Theta} ||G(t, \theta)|| \;\leq\; \frac{1}{(1 + Y_{t-1}^2)} \left|\left|(1, 2|\varepsilon_t Y_{t-1}| + 2n^{-\frac{1}{2\gamma}}Y_{t-1}^2, Y_{t-1}^2)^T\right|\right| \leq C\left(1 + \frac{|\varepsilon_t Y_{t-1}|}{1 + Y_{t-1}^2}\right).$$

Since $\mathrm{E}[|\varepsilon_t Y_{t-1}|/(1+Y_{t-1}^2)]^{2\alpha_0}<\infty$, we can show that

$$\sup_{1\le t\le n}\sup_{\theta\in\Theta}||G(t,\theta)||=o_p(n^{\frac{1}{2\gamma}}).$$

Further, by Lemma 4.1 of Chan, Li, Peng and Zhang (2013[63]), we have

$$\sup_{1\le t\le n}\sup_{\theta\in\Theta}|X_{t1}(\theta,\alpha_0)|=o_p(n^{\frac{1}{2\gamma}}).$$

Hence Lemma 5.4 follows from the above two equations. □

Proof of Theorem 7. It follows from Theorem 1 of Chan and Peng (2005[103]) that there exists a positive matrix Ω_1 such that

$$\sqrt{n}\{\hat{\theta}-\theta_0\}\xrightarrow{d}N(0,\Omega_1). \tag{B.48}$$

Like the proofs in Chan, Li, Peng and Zhang (2013[63]), we can show that

$$\begin{aligned}
&\sqrt{n}(\hat{\alpha}-\alpha_0)\\
=\;&\sqrt{n}(\hat{\omega}-\omega_0)\kappa_0^{-1}\mathrm{E}\Big[\frac{\sqrt{b_0}(\alpha_0|a_0+\sqrt{b_0}\varepsilon_2|^{\alpha_0-1}\mathrm{sgn}(a_0+\sqrt{b_0}\varepsilon_2))\varepsilon_2}{2(\omega_0+b_0Y_1^2)}\Big]\\
&+\sqrt{n}(\hat{a}-a_0)\kappa_0^{-1}\mathrm{E}\Big[(\alpha_0|a_0+\sqrt{b_0}\varepsilon_2|^{\alpha_0-1}\mathrm{sgn}(a_0+\sqrt{b_0}\varepsilon_2))\Big(\frac{\sqrt{b_0}Y_1}{\sqrt{\omega_0+b_0Y_1^2}}-1\Big)\Big]\\
&+\sqrt{n}(\hat{b}-b_0)\kappa_0^{-1}\mathrm{E}\Big[(\alpha_0|a_0+\sqrt{b_0}\varepsilon_2|^{\alpha_0-1}\mathrm{sgn}(a_0+\sqrt{b_0}\varepsilon_2))[\frac{\sqrt{b_0}\varepsilon_2Y_1^2}{2(\omega_0+b_0Y_1^2)}-\frac{\varepsilon_2}{2\sqrt{b_0}}]\Big]\\
&-\frac{1}{\kappa_0\sqrt{n}}\sum_{t=1}^{n}(|a_0+\sqrt{b_0}\varepsilon_t|^{\alpha_0}-\mathrm{E}|a_0+\sqrt{b_0}\varepsilon_1|^{\alpha_0})+o_p(1),
\end{aligned} \tag{B.49}$$

where $\kappa_0=\mathrm{E}(|a_0+\sqrt{b_0}\varepsilon_1|^{\alpha_0}\log|a_0+\sqrt{b_0}\varepsilon_1|)$. Thus, Theorem 7 follows from (B.48), (B.49) and the CLT for martingales. □

Lemma 10. *Under conditions of Theorem 2.3, we have, as $n \to \infty$*

$$(a)\ \sup_{\theta\in\Theta} || \frac{1}{n} \sum_{t=1}^{n} X_t(\theta,\alpha_0)X_t^T(\theta,\alpha_0) - \mathrm{E}\{X_1(\theta_0,\alpha_0)X_1^T(\theta_0,\alpha_0)\}|| = o_p(1);$$

$$(b)\ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t(\theta_0,\alpha_0) \xrightarrow{d} \mathrm{N}\Big(0, \mathrm{E}\{X_1(\theta_0,\alpha_0)X_1^T(\theta_0,\alpha_0)\}\Big).$$

Proof. Conclusion (a) can be proved in a way similar to the proof of Lemma 5.3.

Conclusion (b) follows from the CLT for martingales by noting that

$$\begin{aligned} X_t(\theta_0,\alpha_0) &= \Big(|a_0 + \sqrt{b_0}\varepsilon_t|^{\alpha_0} - 1,\ -\tfrac{\mathrm{sgn}(\varepsilon_t^2-1)}{1+Y_{t-1}^2}, \\ &\quad -\frac{2(\omega_0 + b_0 Y_{t-1}^2)^{\frac{1}{2}} Y_{t-1}\mathrm{sgn}(\varepsilon_t^2-1)\varepsilon_t}{1+Y_{t-1}^2}, \frac{Y_{t-1}^2\mathrm{sgn}(\varepsilon_t^2-1)}{1+Y_{t-1}^2}\Big)^T \end{aligned}$$

is a martingale difference sequence. □

Lemma 11. *Under conditions of Theorem 7,*

$$\begin{aligned} \sup_{\theta\in\Theta} \Big\| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \Big\{ &\left[sgn\left(\varepsilon_t^2(\omega,a,b)\right) - sgn\left(\varepsilon_t^2 - 1\right)\right] H(t,\theta) \\ &+ 2(\theta-\theta_0)^T f_{\varepsilon_1^2}(1) H(t,\theta_0) H^T(t,\theta_0)\Big\}\Big\| = o_p(1), \end{aligned}$$

where $f_{\varepsilon_1^2}(\cdot)$ denotes the density of ε_1^2.

Proof. This lemma can be proved in a way similar to the proof of Lemma 5.3, hence we omit the details. □

Proof of Theorem 2.4. Theorem 2.4 can be shown similar to the proof of Theorem 2.2 by using Lemmas 5.5 and 5.6. □