

THE UNIVERSITY OF CHICAGO

INFERENCE FOR TIME SERIES IN RISK MEASURE AND REGRESSION

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF STATISTICS

BY
WENTAO HU

CHICAGO, ILLINOIS

AUGUST 2025

To my family

TABLE OF CONTENTS

LIST OF FIGURES	vi
LIST OF TABLES	vii
ACKNOWLEDGMENTS	ix
ABSTRACT	x
1 INTRODUCTION	1
1.1 Risk measures	1
1.2 Regression with multivariate time-varying random coefficients	2
1.3 High confidence level inference	4
2 ASYMPTOTICS FOR EXPECTILE ESTIMATORS OF STATIONARY PROCESSES	6
2.1 Introduction	6
2.2 Asymptotic Theory for Expectile	10
2.2.1 Functional dependence measures	11
2.2.2 CLT for expectile estimators at a single level	13
2.2.3 Weak convergence for expectile processes	14
2.2.4 Expectile estimates for long-memory processes	16
2.2.5 Application to G/G/1 Queues	18
2.3 Asymptotic Theory for Expected Shortfall	19
2.3.1 ES for short range dependent processes	19
2.3.2 ES for long-memory processes	21
2.4 Simulation	22
2.4.1 Expectile Estimation for SRD Processes	22
2.4.2 Expectile Estimation for LRD Processes	38
2.5 Application	42
2.6 Conclusion	44
3 SIMULTANEOUS INFERENCE OF REGRESSION WITH MULTIVARIATE TIME-VARYING RANDOM COEFFICIENTS	45
3.1 Introduction	45
3.2 Methodology	48
3.2.1 Estimation	49
3.2.2 Construction of <i>SCR</i>	50
3.2.3 Multiplier bootstrap	54
3.3 Simulation	55
3.4 Application: Currency Market Efficiency	59
3.5 Conclusion	62

4	HIGH CONFIDENCE LEVEL INFERENCE FOR QUANTILE REGRESSION	67
4.1	Introduction	67
4.2	Main Results	70
4.3	Application: Tax Progressivity	71
4.3.1	Data	71
4.3.2	Empirical framework for tax progressivity	72
4.3.3	Inference for dynamic quantile progressivity	74
4.4	Conclusion	76
	REFERENCES	78
A	APPENDIX: SUPPLEMENTARY MATERIALS FOR CHAPTER 2	94
A.1	Additional Simulation	94
A.1.1	Simulations on Queueing processes	94
A.1.2	Simulation on High Confidence Level Inference	101
A.2	Technical Proofs	101
A.2.1	Proof of Theorem 2.2.1	101
A.2.2	Proof of Corollary 2.2.1.1	110
A.2.3	Proof of Theorem 2.2.2	110
A.2.4	Proof of Theorem 2.2.4	115
A.2.5	Proof of Lemma 2.3.3	117
A.2.6	Proof of Theorem 2.3.4	118
A.2.7	Proof of Theorem 2.3.5	119
A.2.8	Proof of Theorem 2.3.6	120
B	APPENDIX : SUPPLEMENTARY MATERIALS FOR CHAPTER 3	121
B.1	Factor pricing model	121
B.2	Technical Proofs	125
B.2.1	Lemmas	125
B.2.2	Proof of Theorem 3.2.1	130
B.2.3	Proof of Theorem 3.2.2	135
B.2.4	Proof of Theorem 3.2.3	138
C	APPENDIX: SUPPLEMENTARY MATERIALS FOR CHAPTER 4	142
C.1	Additional Figures	142
C.2	Technical Proofs	142

LIST OF FIGURES

2.1	SCB in blue vs PCB in yellow with estimated expectiles in black for the AR(1) process with $t(5)$ innovations	37
2.2	Plot of differences $\hat{\theta}_\alpha - \theta_\alpha$, $0.1 \leq \alpha \leq 0.9$ (black) vs $\bar{X}_n$ (red)	38
2.3	SCBs (blue) vs PCBs (yellow) for LRD linear processes with $\beta = 0.55$ and 0.7	41
2.4	Network traffic measured in microseconds, starting at 14:01:00 UTC lasting 11.46 seconds on January 17, 2019	41
2.5	SCBs (blue) vs PCBs (yellow) for network traffic data with 95% coverage probability in the left figure and 99% in the right figure	43
3.1	Intercept (<i>i.e.</i> $\alpha(\cdot)$) and UIP coefficient (<i>i.e.</i> $\beta(\cdot)$) in the relative money growth and the relative output growth for Swiss Franc (CHF)–US Dollar (USD) ; The bandwidth is obtained through the generalized cross-validation (<i>GCV</i>).	64
3.2	The <i>solid</i> curve is the local-linear regression estimate of $\alpha(\cdot)$ in relative money growth for CHF/USD . The <i>dashed</i> band is the 95% <i>SCR</i> of $\alpha(\cdot)$ and the dotted horizontal line is $H_0 : \alpha(\cdot) = 0$. The bandwidth is chosen by <i>GCV</i> . Here (a)–(c) are 25th, 50th and 75th percentile of output growth, respectively.	64
3.3	The <i>solid</i> curve is the local-linear regression estimate of $\alpha(\cdot)$ in relative output growth for CHF/USD . The <i>dashed</i> band is the 95% <i>SCR</i> of $\alpha(\cdot)$ and the dotted horizontal line is $H_0 : \alpha(\cdot) = 0$. The bandwidth is chosen by <i>GCV</i> . Here (a)–(c) are 25th, 50th and 75th percentile of money growth, respectively.	65
3.4	The <i>solid</i> curve is the local-linear regression estimate of $\beta(\cdot)$ in relative money growth for CHF/USD . The <i>dashed</i> band is the 95% <i>SCR</i> of $\beta(\cdot)$ and the dotted horizontal line is $H_0 : \beta(\cdot) = 1$. The bandwidth is chosen by <i>GCV</i> . Here (a)–(c) are 25th, 50th and 75th percentile of output growth, respectively.	65
3.5	The <i>solid</i> curve is the local-linear regression estimate of $\beta(\cdot)$ in relative outout growth for CHF/USD . The <i>dashed</i> band is the 95% <i>SCR</i> of $\beta(\cdot)$ and the dotted horizontal line is $H_0 : \beta(\cdot) = 1$. The bandwidth is chosen by <i>GCV</i> . Here (a)–(c) are 25th, 50th and 75th percentile of money growth, respectively.	66
4.1	Yearly Progressivity Estimates by Income Quantiles	73
4.2	Progressivity Estimates and 95% Simultaneous Confidence Intervals	75
B.1	Pricing error and the beta coefficient of a <i>CAPM</i> in the market portfolio's dividend yield and the de-trended T-bill yield; The bandwidth is obtained through the generalized cross-validation (<i>GCV</i>).	123
B.2	The <i>solid</i> curve is the local-linear regression estimate of the pricing error $\alpha(\cdot)$. The <i>dashed</i> band is the 95% <i>SCR</i> of $\alpha(\cdot)$ and the dotted horizontal line is the null $H_0 : \alpha(\cdot) = 0$. The bandwidth is chosen by <i>GCV</i>	124
C.1	Progressivity Estimates and Simultaneous Confidence Intervals	142

LIST OF TABLES

2.1	Empirical coverage probabilities of two-sided 95% CIs of expectiles for the AR(1) process with $t(3)$ innovations using the d -fold method with sample size n and expectile level α	24
2.2	Halfwidths (multiplied by $\sqrt{n}$) of two-sided 95% CIs of expectiles for the AR(1) process with innovations $t(3)$ using the d -fold method with sample size n and expectile level α	25
2.3	Empirical coverage probabilities of two-sided 95% CIs of expectiles for the TAR(1) process with innovations $N(0, 1)$ using the d -fold method with sample size n and expectile level α	27
2.4	Halfwidths (multiplied by $\sqrt{n}$) of two-sided 95% CIs of expectiles for the TAR(1) process with innovations $N(0, 1)$ using the d -fold method with sample size n and expectile level α	28
2.5	Empirical coverage probabilities of two-sided 95% CIs of expectiles for the TAR(1) process with innovations $t(3)$ using the d -fold method with sample size n and expectile level α	30
2.6	Halfwidths (multiplied by $\sqrt{n}$) of two-sided 95% CIs of expectiles for the TAR(1) process with innovations $t(3)$ using the d -fold method with sample size n and expectile level α	31
2.7	Grand expectiles of two-sided 95% expectile CIs using the d -fold method with sample size n and expectile level α	33
2.8	Relative precision (%) of two-sided 95% expectile CIs using the d -fold method with sample size n and expectile level α	34
2.9	Empirical coverage probabilities of two-sided 95% expectile CIs using the d -fold method with sample size n and expectile level α	35
2.10	Empirical coverage probability of two-sided 95% CIs of expectiles for the LRD linear process with $\beta = 0.7$ and innovations $N(0, 1)$ using the d -fold method with sample size n and expectile level α	39
2.11	Uniform coverage probabilities for AR(1) and LRD processes with $N(0, 1)$ and $t(3)$ innovations	39
3.1	95% coverage probability of SCR for $\alpha(\cdot, \cdot)$ and $\beta(\cdot, \cdot)$, where Z_{1i} and Z_{2i} follow a VAR(1)	58
A.1	Empirical coverage probabilities of two-sided 95% CIs of expectiles for the M/M/1 process using the d -fold method with sample size n and expectile level α	95
A.2	Halfwidths (multiplied by $\sqrt{n}$) of two-sided 95% CIs of expectiles for the M/M/1 process using the d -fold method with sample size n and expectile level α	96
A.3	Empirical coverage probabilities of two-sided 95% CIs of expectiles for the M/H ₂ /1 process using the d -fold method with sample size n and expectile level α	97
A.4	Halfwidths (multiplied by $\sqrt{n}$) of two-sided 95% CIs of expectiles for the M/H ₂ /1 process using the d -fold method with sample size n and expectile level α	98

A.5	Empirical coverage probabilities of two-sided 95% CIs of expectiles for the M/G/1 process with lognormal service time using the d -fold method with sample size n and expectile level α	99
A.6	Halfwidths (multiplied by $\sqrt{n}$) of two-sided 95% CIs of expectiles for the M/G/1 process with lognormal service time using the d -fold method with sample size n and expectile level α	100
A.7	Empirical coverage probabilities of two-sided 99% expectile CIs for the AR(1) process with $t(3)$ innovations using the d -fold method with sample size n and expectile level α with 10,000 replications	102
A.8	Halfwidths (multiplied by $\sqrt{n}$) of two-sided 99% expectile CIs for the AR(1) process with innovations $t(3)$ using the d -fold method with sample size n and expectile level α with 10,000 replications	103
A.9	Empirical coverage probabilities of two-sided 99% CIs of expectiles for the TAR(1) process with innovations $\mathbf{N}(0, 1)$ using the d -fold method with sample size n and expectile level α with 10,000 replications	104
A.10	Halfwidths (multiplied by $\sqrt{n}$) of two-sided 99% CIs of expectiles for the TAR(1) process with innovations $\mathbf{N}(0, 1)$ using the d -fold method with sample size n and expectile level α with 10,000 replications	105
A.11	Empirical coverage probabilities of two-sided 99% CIs of expectiles for the TAR(1) process with innovations $t(3)$ using the d -fold method with sample size n and expectile level α with 10,000 replications	106
A.12	Halfwidths (multiplied by $\sqrt{n}$) of two-sided 99% CIs of expectiles for the TAR(1) process with innovations $t(3)$ using the d -fold method with sample size n and expectile level α with 10,000 replications	107

ACKNOWLEDGMENTS

Above all, I am deeply and profoundly grateful to my advisor, Professor Wei Biao Wu. His guidance and support have shaped every stage of my PhD journey. From my very first class at the University of Chicago—Stat 304—I was inspired by his elegant and rigorous introduction to probability and statistics, from foundational concepts to advanced theories. He deepened my understanding of the logic and motivation behind statistical results and taught me that academic research should be both natural and rigorous.

Throughout these years, working under Prof. Wu’s mentorship has not only advanced my understanding of time series research, but also refined my academic taste and strengthened my confidence as a researcher. I am proud of the work I have done and remain optimistic about the challenges that lie ahead. This belief is deeply rooted in the academic training and values I have received from him and will continue to guide my future research. Without him, none of this would have been possible.

I would also like to sincerely thank my committee members, Professor Chao Gao and Professor Mei Wang, for their insightful comments and crucial support throughout my proposal and dissertation process. I am especially grateful to my collaborators, Professor Kun Ho Kim and Dr. Jiheum Yeon, who introduced me to cutting-edge research topics in financial economics and helped bring theoretical models into meaningful application—particularly in Chapters 3 and 4. I also appreciate the collaboration with Dr. Rui Ren and Dr. Yifu Wang, whose perspectives in management science broadened the real-world depth and demand for risk measures, as discussed in Chapter 2.

Finally, I thank my family. To my parents, for their constant love and unwavering support across the distance. And to my wife, Linwei Gao, for her patience, encouragement, and endless care—your love is always my anchor and my strength. Words fall short of capturing the full depth of my gratitude. All of these pieces—each act of support, each shared insight—have made my PhD journey not only possible but profoundly meaningful.

ABSTRACT

This dissertation mainly studies the statistical inference on expectile, time-varying regression coefficients and quantile regression coefficients for time series. For the first problem, We first present a central limit theorem at a single level and then develop a simultaneous inference theory over various levels under the short-range dependent condition. We also derives a weak convergence theory for expectile estimators of long-memory processes, and reveals the surprising fact that pointwise and simultaneous confidence bands are asymptotically the same.

For the second problem, we introduce a new methodology to conduct simultaneous inference of the nonparametric coefficient functions in a linear time series regression model where the nonparametric components are unknown multivariate functions of an observable random vector. In particular, we construct simultaneous confidence regions (SCR) for the coefficient functions by extending the high-dimensional Gaussian approximation to time-dependent processes.

For the third problem, we establish the Gaussian approximation theory for quantile regression coefficients on high confidence level inference. Based on the quantile regression model, we investigates the dynamics of U.S. tax progressivity from 1978 to 2022 by conducting year-by-year quantile regression and simultaneous inference on progressivity estimates. We can find the constancy of the progressivity is rejected in the 1980s with a gradual decline and during recessions with jumps in the progressivity.

CHAPTER 1

INTRODUCTION

Time-dependent data arises naturally in a wide range of fields, including economics, finance, management. The presence of temporal dependence often introduces substantial challenges in both theory and practice. This dissertation investigates three distinct interesting problems that emerge when conducting statistical analysis under time dependence.

1.1 Risk measures

Risk measurement plays a foundational role in decision sciences. A well-designed risk measure provides a formal and interpretable way to quantify potential losses, support portfolio construction, and guide decision-making under uncertainty. Among the most prominent risk measures are Value-at-Risk (VaR), Expected Shortfall (ES), and expectiles. Statistically, they are either quantile-based or moment-based.

Let X be a random variable with $\mathbb{E}|X| < \infty$ and F be the cumulative distribution function of X . Then for risk level $\alpha \in (0, 1)$, Value-at-Risk, Expected Shortfall and expectile can be respectively defined as

$$\begin{aligned}\text{VaR}_\alpha(X) &= -\inf\{x \in \mathbb{R} : F(x) > \alpha\}, \\ \text{ES}_\alpha(X) &= \frac{1}{\alpha} \int_0^\alpha \text{VaR}_\gamma(X) d\gamma, \\ \text{Expectile}_\alpha(X) &= \underset{\theta}{\operatorname{argmin}} \mathbb{E}[\alpha(X - \theta)^2 \mathbb{1}_{X \geq \theta} + (1 - \alpha)(X - \theta)^2 \mathbb{1}_{X < \theta}].\end{aligned}$$

As we can see from the definitions above, VaR, based on the quantile of distribution, can indicate the loss boundary at certain risk level. It has been widely adopted in regulatory and industry contexts due to its intuitive interpretation. However, VaR suffers from several known limitations: it lacks subadditivity, fails to capture tail severity, and is not elicitable.

ES addresses some of these concerns by considering the average loss beyond the VaR threshold. It is coherent and captures tail risk more comprehensively. Yet, ES is not elicitable as a standalone risk measure, which complicates its statistical implementation and inference.

Expectiles have recently gained significant attention as a compelling alternative. Defined through asymmetric least squares, expectiles incorporate both tail probability and tail magnitude, offering a more sensitive assessment of risk. Importantly, they are elicitable and law-invariant coherent. However, expectile is not comonotonic additive. In one sentence, currently, there does not exist a perfect standard measure. Therefore, further investigation into risk measures remains highly valuable.

Although VaR has been extensively studied in the statistical literature, the development of rigorous theory for ES and expectiles remains relatively limited. Most existing results for ES and expectiles are established under i.i.d. assumptions or within specific parametric models. This gap motivates the investigation in Chapter 2 of this dissertation. We study the asymptotic behavior of expectile estimators under various types of dependence structures. In particular, we consider short-range dependent (SRD) processes, long-range dependent (LRD) processes, and steady-state simulations, all of which frequently arise in applied settings. We develop central limit theorems, uniform weak convergence results, and construct confidence intervals and confidence bands suitable for such contexts. These contributions enhance the theoretical foundation for using expectiles in applied risk analysis and broaden their practical applicability.

1.2 Regression with multivariate time-varying random coefficients

Consider the very classical linear model:

$$Y_i = \boldsymbol{\beta}^\top X_i + \varepsilon_i, \quad i = 1, \dots, n,$$

where Y_i are responses, $\beta \in \mathbb{R}^p$ is the p –dimensional coefficient vector, X_i are p –dimensional covariates and ε_i are noise. With linear assumption, we can conduct estimation, inference and interpretation on various datasets. And many advanced and powerful results have been derived during past decades. However, linear assumption does not always hold. Then a more general framework has been studied:

$$Y_i = \alpha(X_i) + \varepsilon_i, \quad i = 1, \dots, n,$$

where $\alpha(\cdot)$ is a well-defined coefficient function and the variance of noise can also depend on X_i .

The classical nonparametric regression framework offers a flexible approach to modeling the relationship between a response and covariates, without assuming a specific functional form. However, its applicability is fundamentally limited by the curse of dimensionality, which renders it practically infeasible to model interactions between the response and more than a small number of covariates. In fact, asymptotic theory for kernel-based methods with more than two covariates is scarce and often relies on strong assumptions that limit its generalizability.

On the other hand, many econometric models of substantive practical relevance—such as asset pricing models, monetary policy rules—are based on linear structures whose coefficients are directly linked to interpretable quantities. For instance, the slope coefficient in a pricing model may represent the risk premium associated with a specific factor. In such settings, it is of primary interest not only to model these coefficients as functions of latent state variables but also to conduct formal inference on their functional form, as these functions encode economic mechanisms or policy regimes. The semi-parametric varying coefficient model, which allows the coefficients to vary nonparametrically with respect to some state process, thus provides a natural and interpretable compromise between the flexibility of nonparametric models and the interpretability of linear parametric structures.

Motivated by these considerations, in Chapter 3, we study the semi-parametric time series regression model:

$$Y_i = \beta(Z_i)^\top X_i + \sigma(Z_i)\varepsilon_i, \quad i = 1, \dots, n.$$

And we develop a comprehensive framework for conducting simultaneous inference on both the coefficient functions $\beta(\cdot)$ and the conditional variance function $\sigma(\cdot)$ in the presence of time series dependence. Specifically, we construct uniform confidence bands for all $\beta(\cdot)$ and $\sigma(\cdot)$, enabling joint hypothesis testing and specification analysis. To facilitate the practical implementation of the proposed inference procedures, we further develop a bootstrap-based method that ensures valid finite-sample performance.

1.3 High confidence level inference

In many empirical applications, statistical inference is typically conducted with confidence levels such as 90% or 95%. However, there is a growing set of problems in which researchers or policymakers require extremely high-confidence level inference — for example, confidence levels of 99.9% or higher. This need arises in various contexts: when a rare event is of major concern, such as in public safety or regulatory policy; or when Bonferroni-type adjustments are applied due to multiple testing, requiring much stricter control of Type I error. In such high-confidence regimes, standard inferential procedures—such as the Central Limit Theorem (CLT)—may no longer yield reliable results in terms of relative error. In other words, conventional approximations can become untrustworthy exactly in the tail regions where high confidence is needed most.

A concrete example of such a high-confidence setting arises in the study of tax progressivity and its long-run economic effects. To estimate tax progressivity, we typically rely on income data spanning several decades, during which tax policies change discontinuously

across different presidential administrations. One might be tempted to smooth the trend using standard techniques such as kernel regression or sliding window averages to extract long-term trends. However, such smoothing procedures can severely distort the causal interpretation of the data: they blur the boundaries between distinct policy regimes and dilute the sharp effects associated with specific tax reforms. In this context, high-confidence inference methods, which directly target the simultaneous tail behavior of estimators, offer a compelling alternative. In Chapter 4, we will investigate the tax progressivity problem in detail and develop high confidence inference framework for quantile regression.

CHAPTER 2

ASYMPTOTICS FOR EXPECTILE ESTIMATORS OF STATIONARY PROCESSES

2.1 Introduction

Expectiles are widely used in a variety of applications in management science and operation research, including risk management, portfolio evaluation and optimization, and decision making under uncertainties. A proper risk measure is fundamental to risk management. Value-at-Risk (VaR) is a popular risk indicator, which is the quantile with a given tail probability. However, it depends only on the relative frequency of more extreme loss rather than the magnitude of the loss. In contrast, Expectile based VaR (EVaR), i.e., the expectile with a given distribution, is an attractive alternative [Kuan et al., 2009]. It is determined by tail expectations rather than tail probabilities, and balances between expected margin shortfall and opportunity cost. EVaR is proved to be an elicitable risk measure [Ziegel, 2016], and has a close relationship with VaR and expected shortfall (ES). The connection between EVaR and VaR is discussed by Jones [1994], which shows that expectiles of a distribution F correspond to quantiles of a distribution G transformed from the original distribution F . Yao and Tong [1996] study the one-to-one mapping between expectile and quantile in a nonparametric setting. Inspired by the latter relation, Waltrup et al. [2015] conduct numerical analysis of empirical quantiles and expectile-based quantile estimators. Also, Taylor [2008] shows that ES can be calculated by EVaR. Ren et al. [2022] propose a systemic risk indicator Financial Risk Meter (FRM) which is based on expectiles.

For portfolio evaluation and optimization, when using VaR as the evaluation criterion, it is possible for two portfolios with entirely different tail structures to exhibit the same quantile. In comparison, expectiles are more tail-sensitive and can avoid the issue accordingly. While quantiles suffer from the incoherence problem and expected shortfall lacks robust-

ness and elicibility, expectiles also have downsides – specifically, they are neither robust nor comonotonically additive [Emmer et al., 2015]. Nevertheless, Ziegel [2016] shows that expectiles are indeed the only law-invariant and coherent elicitable risk measures, which is critical for portfolio evaluation as sub-additivity implies that diversification never increases risk. Expectiles can also be used to optimize portfolios by balancing risk and return. This approach is particularly useful in volatile scenarios where traditional risk indicators, e.g., variance and standard deviation, fail to capture the underlying risks. For example, Giacometti et al. [2021] introduce the asymmetric expectile loss function to account for tail risk in large portfolio allocation. Bellini et al. [2021] propose three methods for constructing risk parity portfolios with respect to expectiles. Zaeviski and Nedeltchev [2023] propose an expectile-based risk measure and demonstrate its significant potential for inclusion in Basel IV, which is a set of regulatory capital standards and latest iteration of a series of international banking agreements aiming to enhance the stability and standardization of the global banking system.

Expectiles serve as a valuable tool in decision making under uncertainty, as they allow decision makers to consider a range of potential outcomes, and provide a shortfall risk measure [Delage et al., 2019], especially in the presence of heavy-tailed distributions. For example, in project management, expectiles are used to measure the risk of a project. By considering different expectiles, such as the 0.05 expectile or the 0.95 expectile, decision makers can determine the worst and best-case scenarios for a project, and make decisions accordingly. Expectile models are also useful in categorical data analysis (e.g., analysis of survey data with sub-populations labelled as areas) as they can handle a number of categories, whereas many traditional methods, such as multinomial logistic mixed models, are limited to only a few categories. They also provide a useful alternative to mixed model approaches since they do not require specific parametric assumptions about random effects [Dawber et al., 2022]. Another example is geoadditive expectile regression [Sobotka and

Kneib, 2012, Sobotka et al., 2013], where the specification allows for nonlinear effects and the spatial distribution of survey data simultaneously.

Due to the above-mentioned theoretical and practical advantages, expectiles have drawn plenty of attention in management science, statistics, econometrics and operations research. However, the attributes of expectile estimates are not well-studied compared with quantiles and ES. For instance, Embrechts et al. [2022] study the robustness of risk-measurement-related optimization problems given special attention to VaR and ES but only with a very limited discussion on expectiles. Wang and Zitikis [2021] omit the discussion on expectiles for the necessary and sufficient condition of an inequality (inequality (1) in their paper) due to the complicated structure of expectiles.

The main goal of this paper is to study the asymptotic properties of expectile estimators for stationary processes, which are crucial for understanding the reliability and accuracy of expectile estimators in large samples. There is an increasing amount of work concerning this topic, e.g., Holzmann and Klar [2016] study the asymptotic properties of the sample expectiles for independent and identically distributed (i.i.d.) random variables. Philipps [2022a,b] develop the weak law of large numbers and uniform ergodic theorem for expectiles. Philipps [2023] investigates into the quasi-maximum likelihood estimator for conditional expectiles. Most of the previous work postulates that the errors are independent. However, the assumption of independence can be a notable limitation when dealing with temporally dependent data. We also avoid strong mixing conditions and their variants, because they are difficult to check [Bradley, 2005a]. Geometric-moment contraction (GMC) condition proposed by Wu [2005a] is broadly applicable and can be checked empirically as illustrated by Alexopoulos et al. [2019]. Therefore, in this paper, we adopt the short-range dependence (SRD) structure; see Equation (2.9), which is widely applicable including a vast class of processes as shown in Section 2.2.1, and easy to work with since it is directly related to the data-generating mechanism [Wu, 2005b].

We first present a central limit theorem (CLT) for expectile estimators at a single level α under the above-mentioned SRD condition. Accordingly, one can calculate confidence intervals (CIs) by using t -distribution. In practice, the constructed CIs have a quite accurate coverage probability and the length of the CIs is relatively stable. This paper then studies simultaneous inference of expectile estimators over $\alpha \in [\alpha_1, \alpha_2]$ with $0 < \alpha_1 < \alpha_2 < 1$. For i.i.d. random variables with finite second moment and continuous distribution functions, Holzmann and Klar [2016] present a weak convergence theory for the expectile process. Their theory has limited applicability in our setting, as it relies on the independence assumption. In this paper, we shall develop a weak convergence theory using the framework of functional dependence measure with some extra mild conditions. With our weak convergence theory, we can construct simultaneous confidence bands (SCBs) and pointwise confidence bands (PCBs) for various stationary processes, such as Autoregressive (AR) and Threshold Autoregressive (TAR) models.

The paper also studies the asymptotic properties of long-memory or long-range dependent (LRD) processes. In such processes, the auto-correlation function decays fairly slowly and they have been widely used in economics [Selmi and Hachicha, 2015], finance [Linn and Tay, 2007, Shi and Yu, 2023], environment sciences [Pan and Chen, 2008, Yuan et al., 2022], internet traffic analysis [Resnick and Samorodnitsky, 2000, Yao, 2003, Sohn and Lee, 2004, Çağlar, 2004, De Lima and de Almeida Amazonas, 2013], risk management [Ho and Liu, 2010, Pankratova et al., 2017] among others. We develop a corresponding weak convergence theory and implement CB simulation for such processes. It reveals the very surprising and interesting fact that for long-memory processes, PCBs and SCBs are asymptotically the same.

We also establish the asymptotic validity of CIs and CBs for expectiles in the steady-state setting. Steady-state simulations are crucial in evaluating the long-run performance or risk of complex systems, and they are broadly applicable in analyzing job cycle times in manufac-

turing systems, message transmission times in communication networks and portfolios in the financial industry [Alexopoulos et al., 2019]. This paper provides CI and CB estimation for steady-state expectiles of simulation generated processes, such as the waiting-time process arising from a steady-state G/G/1 queueing system. These findings are valuable since the SRD condition enables specific output-analysis procedures to work reliably when applied to a broad class of output processes that are generated by simulations in a steady-state operation.

The rest of the paper is organized as follows. In Section 2.2, we introduce functional dependence measures and derive asymptotic properties for expectile estimators under different scenarios. Section 2.4 presents simulation studies. Section 2.5 provides an application to an internet traffic data. Section 2.6 concludes.

2.2 Asymptotic Theory for Expectile

Let $(X_t)_{t \in \mathbb{Z}}$ be a stationary process. For $0 < \alpha < 1$, define

$$g_\alpha(x) = \begin{cases} \alpha x, & x \geq 0, \\ (1 - \alpha)x, & x < 0. \end{cases} \quad (2.1)$$

Then the expectile θ_α is the solution to $\mathbb{E}g_\alpha(X_t - \theta_\alpha) = 0$. Equivalently, θ_α is the minimizer of the asymmetric quadratic loss function [Newey and Powell, 1987]:

$$\theta_\alpha = \operatorname{argmin}_{\theta \in \mathbb{R}} \mathbb{E}[\rho_\alpha(X_t - \theta) - \rho_\alpha(X_t)], \text{ where } \rho_\alpha(x) = \begin{cases} \alpha x^2/2, & x \geq 0, \\ (1 - \alpha)x^2/2, & x < 0. \end{cases}$$

Given the observed data $X_1, \dots, X_n$, the sample expectile $\hat{\theta}_\alpha = \hat{\theta}_\alpha(X_1, \dots, X_n)$ satisfies

$$Q_{\alpha,n}(\hat{\theta}_\alpha) = 0, \text{ where } Q_{\alpha,n}(\theta) := \frac{1}{n} \sum_{t=1}^n g_\alpha(X_t - \theta). \quad (2.2)$$

Note that the solution $\hat{\theta}_\alpha$ is unique since $g_\alpha(\cdot)$ is continuous and strictly increasing. In the special case when $\alpha = 1/2$, then $\theta_\alpha = \mathbb{E}X_t$ and $\hat{\theta}_\alpha = \bar{X}_n$, which is the mean and sample mean, respectively.

The primary goal of this paper is to establish an asymptotic theory for the estimate $\hat{\theta}_\alpha$ for stationary processes. For i.i.d. random variables, Holzmann and Klar [2016] present a systematic asymptotic theory; see also Tran et al. [2019]. It is substantially more challenging to develop an asymptotic theory for stationary processes, partially because expectile estimators are nonlinear functions of the data. Some recent developments are made in Zwingmann and Holzmann [2020], Krättschmer and Zähle [2017], Coeurjolly and Kortas [2012] among others.

2.2.1 Functional dependence measures

In this paper we shall work with the stationary causal process of the form

$$X_t = s(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots), \quad (2.3)$$

where ε_t , $t \in \mathbb{Z}$, are i.i.d random variables and s is a measurable function such that X_t is well-defined. In (2.3), X_t is causal in the sense that X_t does not depend on the future innovations ε_j , $j > t$. The causality is a reasonable assumption in the study of time series. The class of processes denoted by (2.3) is vast, including the linear process

$$X_t = \sum_{i=0}^{\infty} a_i \varepsilon_{t-i}, \quad (2.4)$$

where a_i are real coefficients such that X_t exists, and the Volterra process [Giannakis and Serpedin, 2001]

$$X_t = \sum_{0 \leq i_1 < \dots < i_k} a_{i_1, \dots, i_k} \varepsilon_{t-i_1} \cdots \varepsilon_{t-i_k}, \quad (2.5)$$

where $a_{i_1, \dots, i_k}$ are real coefficients. The framework (2.3) also includes many time series models [Priestley, 1988, Tong, 1990] such as autoregressive conditional heteroscedastic (ARCH) processes [Engle, 1982] and generalized ARCH (GARCH) processes [Bollerslev, 1986], random coefficient autoregressive (RCA) processes [Nicholls and Quinn, 2012] and TAR processes [Tong, 1990] among others. For example, consider the Markov process

$$X_t = G(X_{t-1}, \varepsilon_t), \quad (2.6)$$

where $G(\cdot, \cdot)$ is a measurable function such that $\mathbb{E}(|G(x_0, \varepsilon_0)|^p) < \infty$, $p > 0$, and

$$\mathbb{E}(L_{\varepsilon_0}^p) < 1, \text{ where } L_{\varepsilon_0} = \sup_{x \neq x'} \frac{|G(x, \varepsilon_0) - G(x', \varepsilon_0)|}{|x - x'|}. \quad (2.7)$$

Under the above condition, iterations of (2.6) lead to a stationary solution (2.3); see Wu and Shao [2004] where a GMC property is obtained. Dinggeç et al. [2022] considered a steady-state G/G/1 queueing system which is of form (2.6) with GMC condition; see Section 2.2.5 for details. In certain cases, form (2.3) is also called the nonlinear Wold representation.

To establish an asymptotic theory for estimates of stationary processes, we shall introduce the functional dependence measure [Wu, 2005b]

$$\delta_{n,p} = \|X_n - X'_n\|_p := \{\mathbb{E}|X_n - X'_n|^p\}^{1/p}, \quad (2.8)$$

where $X'_n = s(\varepsilon_n, \dots, \varepsilon_1, \varepsilon'_0, \varepsilon_{-1}, \varepsilon_{-2}, \dots)$ is a coupled version of X_n . Here $\varepsilon'_i, \varepsilon_j, i, j \in \mathbb{Z}$, are i.i.d. and X'_n is obtained by replacing ε_0 in X_n by the i.i.d. copy ε'_0 . In comparison with the strong mixing conditions that are widely used in the asymptotic theory of random processes [Bradley, 2005b, Doukhan, 2012], our functional dependence measure is usually easy to work with since it is directly related to the data-generating mechanism. Additionally, results developed based on functional dependence measures are optimal or nearly optimal

(see, for example, Wu [2007a], Berkes et al. [2014]).

2.2.2 CLT for expectile estimators at a single level

Theorem 2.2.1 presents a multivariate central limit theorem under a mild SRD condition. Let $\hat{\theta}_{\alpha,d} = \hat{\theta}_{\alpha}(X_{1+(d-1)m}, \dots, X_{dm})$ be the α th expectile estimate based on the block $\{X_{1+(d-1)m}, \dots, X_{dm}\}$. Denote by $\mathbf{I}_d$ the d -dimensional identity matrix.

Theorem 2.2.1. *Let $d \in \mathbb{N}$ be fixed. Let $0 < \alpha < 1$ and assume that the distribution function $F(x) = \mathbb{P}(X_1 \leq x)$ is continuous at θ_{α} . Further assume that the SRD condition*

$$\sum_{t=0}^{\infty} \delta_{t,2} < \infty \quad (2.9)$$

holds. Then as $m \rightarrow \infty$, the joint distribution $\hat{\theta}_{\alpha,d} = \hat{\theta}_{\alpha}(X_{1+(d-1)m}, \dots, X_{dm})$ satisfies

$$\sqrt{m}(\hat{\theta}_{\alpha,1} - \theta_{\alpha}, \hat{\theta}_{\alpha,2} - \theta_{\alpha}, \dots, \hat{\theta}_{\alpha,d} - \theta_{\alpha}) \Rightarrow \frac{\sigma_{\alpha\infty}}{\tau_{\alpha}} \mathbf{N}(0, \mathbf{I}_d), \quad (2.10)$$

where $\tau_{\alpha} = \mathbb{E}g'_{\alpha}(X_1 - \theta_{\alpha})$, and the long-run variance

$$\sigma_{\alpha\infty}^2 = \sum_{t=-\infty}^{\infty} \text{Cov}\{g_{\alpha}(X_1 - \theta_{\alpha}), g_{\alpha}(X_t - \theta_{\alpha})\}. \quad (2.11)$$

Remark 2.2.1. A process satisfying SRD condition (2.9) is also called stable (see Definition 3 in Wu [2005b]). For linear process (2.4), we have $\delta_{t,2} = |a_t| \|\varepsilon_0 - \varepsilon'_0\|$. Then condition (2.9) follows from $\sum_{t=0}^{\infty} |a_t| < \infty$, which is the stability condition for linear filter [Box et al., 2015]. For additional examples, see Wu [2005b], who discussed the stable conditions for Volterra processes, nonlinear transforms of linear processes, nonlinear time series and other models.

As an immediate consequence of Theorem 2.2.1, by the continuous mapping theorem, we have the following corollary, which asserts that we can use t -distribution to perform statistical inference for θ_{α} without the need to estimate the variance $\sigma^2 = \sigma_{\alpha\infty}^2 / \tau_{\alpha}^2$ consistently.

Specifically, given the level $\gamma \in (0, 1)$, one can construct the $(1 - \gamma)$ th confidence interval for θ_α by

$$\bar{\theta}_{\alpha,d} \pm d^{-1/2} \hat{\sigma}_{\alpha,d} t_{d-1,1-\gamma/2}, \quad (2.12)$$

where $t_{d-1,1-\gamma/2}$ is the $(1 - \gamma/2)$ th quantile of the t -distribution t_{d-1} with $d - 1$ degrees of freedom, and $\hat{\sigma}_{\alpha,d}$ and $\bar{\theta}_{\alpha,d}$ is defined in the following Corollary 2.2.1.1. Because d is directly related to the block size m , and thus to the asymptotic normality in (A.4), as well as the half length $d^{-1/2} \hat{\sigma}_{\alpha,d} t_{d-1,1-\gamma/2}$, selecting an appropriate d is crucial for deriving confidence intervals with desired coverage probability and decent half length. We will discuss the choice of d carefully in Section 2.4.1.

Corollary 2.2.1.1. *Under conditions of Theorem 2.2.1, for $\bar{\theta}_{\alpha,d} = \sum_{j=1}^d \hat{\theta}_{\alpha,j}/d$, we have*

$$\frac{\sqrt{d}(\bar{\theta}_{\alpha,d} - \theta_\alpha)}{\hat{\sigma}_{\alpha,d}} \Rightarrow t_{d-1}, \text{ where } \hat{\sigma}_{\alpha,d}^2 = \frac{1}{d-1} \sum_{j=1}^d (\hat{\theta}_{\alpha,j} - \bar{\theta}_{\alpha,d})^2. \quad (2.13)$$

2.2.3 Weak convergence for expectile processes

Theorem 2.2.1 and Corollary 2.2.1.1 provide a CLT for a fixed $\alpha \in (0, 1)$. To perform simultaneous inference of $\hat{\theta}_\alpha$ over $\alpha \in [\alpha_1, \alpha_2]$, where $0 < \alpha_1 < \alpha_2 < 1$, we need to develop a weak convergence theory for the expectile process $\{\sqrt{n}(\hat{\theta}_\alpha - \theta_\alpha), \alpha_1 \leq \alpha \leq \alpha_2\}$. To this end, we shall apply the weak convergence theory in Billingsley [2013] by imposing extra mild conditions. For i.i.d. random variables with finite second moment and continuous distribution functions, Holzmam and Klar [2016] derived a weak convergence theory for the expectile process using the classical empirical processes of independent random variables; see Van der Vaart [2000]. Their theory is not applicable in our setting since their argument heavily depends on the independence assumption. For stationary processes, we need to develop new tools to establish the weak convergence using the framework of functional dependence measure.

Theorem 2.2.2. *Assume that the distribution function $F(\cdot)$ is continuous and (2.9) holds. Assume either (i) the functional dependence measure*

$$\delta_{n,p} = O(n^{-K}), \text{ where } K > p/(p-1) \text{ and } p > 2, \quad (2.14)$$

or (ii) the conditional distribution function $\mathbb{E}(\mathbf{1}_{X_t \leq y} | \mathcal{F}_{t-1})$ is SRD:

$$\sum_{t=0}^{\infty} \sup_y \|\mathcal{P}_0 \mathbf{1}_{X_t \leq y}\| < \infty, \text{ where } \mathcal{P}_k \cdot = \mathbb{E}(\cdot | \mathcal{F}_k) - \mathbb{E}(\cdot | \mathcal{F}_{k-1}). \quad (2.15)$$

Let $0 < \alpha_1 < \alpha_2 < 1$. Then we have the weak convergence

$$\{\sqrt{n}(\hat{\theta}_\alpha - \theta_\alpha), \alpha_1 \leq \alpha \leq \alpha_2\} \Rightarrow \left\{ \frac{W(\alpha)}{\tau_\alpha}, \alpha_1 \leq \alpha \leq \alpha_2 \right\}, \quad (2.16)$$

where $W(\alpha)$ is a mean 0 Gaussian process with the long-run covariance function

$$\mathbb{E}(W(\alpha)W(\alpha')) = \gamma_{\alpha\alpha'\infty} \text{ where } \gamma_{\alpha\alpha'\infty} = \sum_{t=-\infty}^{\infty} \text{Cov}\{g_\alpha(X_0 - \theta_\alpha), g_{\alpha'}(X_t - \theta_{\alpha'})\}. \quad (2.17)$$

Remark 2.2.2. With the weak convergence property (2.16), we can construct simultaneous confidence bands for the expectile function $\{\theta_\alpha, \alpha_1 \leq \alpha \leq \alpha_2\}$. Let q_γ be the γ th quantile of $\sup_{\alpha_1 \leq \alpha \leq \alpha_2} |W(\alpha)/\tau_\alpha|$, $0 < \gamma < 1$; let $\hat{q}_\gamma$ be an estimate of q_γ . Then the γ th simultaneous confidence band can be constructed as $\hat{\theta}_\alpha \pm \hat{q}_\gamma/\sqrt{n}$, $\alpha_1 \leq \alpha \leq \alpha_2$. To estimate q_γ in practice, we can use the block method; see Section 2.4.1.

Example 2.2.3. We apply Theorem 2.2.2 to the linear process (2.4). Assume that ε_t are i.i.d. with $\mathbb{E}(|\varepsilon_t|^2) < \infty$, and its distribution function F_ε is Lipschitz continuous, namely there exists a constant L_ε such that $|F_\varepsilon(w) - F_\varepsilon(w')| \leq L_\varepsilon|w - w'|$ holds for all w, w' . Let $a_0 = 1$. Then the conditional distribution $\mathbb{E}(\mathbf{1}_{X_t \leq y} | \mathcal{F}_{t-1}) = F_\varepsilon(y - Y_t)$, where $Y_t = \sum_{j=1}^{\infty} a_j \varepsilon_{t-j}$.

Let $Y_{t,0} = Y_t - a_t \varepsilon_0 + a_t \varepsilon'_0$, where ε'_0 and $\varepsilon_t, t \in \mathbb{Z}$, are i.i.d. Note that

$$\|F_\varepsilon(y - Y_t) - F_\varepsilon(y - Y_{t,0})\| \leq L_\varepsilon \|a_t \varepsilon_0 - a_t \varepsilon'_0\| = L_\varepsilon |a_t| \sqrt{2} \|\varepsilon_0\|.$$

Thus (2.15) holds under the classical SRD condition

$$\sum_{t=0}^{\infty} |a_t| < \infty. \quad (2.18)$$

If the above condition is violated, for example $a_t \asymp t^{-\beta}$ with $\beta \in (1/2, 1)$, then the process is LRD. In this case the resulting weak convergence theory has a very different form; see Theorem 2.2.4 for a detailed description.

2.2.4 Expectile estimates for long-memory processes

The next theorem concerns LRD processes. A widely used long-memory model is the fractionally integrated auto-regressive moving average (FARIMA) process

$$\phi(B)(1 - B)^d X_t = \pi(B)\varepsilon_t, \quad (2.19)$$

where $d \in (0, 1/2)$ is the order of the fractional integration, $\phi(B)$ and $\pi(B)$ are auto-regressive and moving average operators, respectively, and ε_t are i.i.d. Note that (2.19) is a special case of the linear process

$$X_t = \sum_{i=0}^{\infty} a_i \varepsilon_{t-i}, \quad (2.20)$$

where ε_t are i.i.d. with mean 0, $a_0 = 1$, $a_n = n^{-\beta} \ell(n)$, $n \geq 1$, where $1/2 < \beta < 1$, and ℓ is a slowly varying function [Bingham et al., 1989], namely for any constant $b > 0$, $\lim_{n \rightarrow \infty} \ell(bn)/\ell(n) = 1$. The linear process (2.20) is a special case of (2.4) with regularly varying coefficients.

It is well-known that (X_i) has the following central limit theorem (see, for example, Avram and Taqqu [1987]):

$$\frac{\bar{X}_n}{n^{1/2-\beta}\ell(n)\sigma} \Rightarrow N(0, c_\beta^2), \text{ where } c_\beta^2 = \int_0^\infty \left(\int_{t-1}^t \max(u, 0)^{-\beta} du \right)^2 dt \quad (2.21)$$

and $\sigma^2 = \mathbb{E}|\varepsilon_i|^2$. Note that, since $1/2 < \beta < 1$, the normalizing sequence $n^{1/2-\beta}\ell(n)$ is larger than $n^{-1/2}$, which is for short-memory processes. Coeurjolly and Kortas [2012] considered expectiles for subordinated Gaussian processes. Here we are working on the more general case of non-Gaussian linear processes. Sample expectiles $\hat{\theta}_\alpha$ for long-memory processes have the attractive property that $\hat{\theta}_\alpha - \theta_\alpha$ can be uniformly approximated by $\bar{X}_n$ in the sense that

$$\sup_{\alpha \in [\alpha_1, \alpha_2]} |\hat{\theta}_\alpha - \theta_\alpha - \bar{X}_n| = o(n^{1/2-\beta}\ell(n)) \quad (2.22)$$

for all $0 < \alpha_1 < \alpha_2 < 1$. See the proof of Theorem 2.2.4 for a derivation of (2.22).

Theorem 2.2.4. *For the LRD process (2.20), assume that $\mathbb{E}|\varepsilon_i|^p < \infty, p > 2$, and f_ε , the probability density function of ε_t , is differentiable and satisfies $\sup_x (f_\varepsilon(x) + |f'_\varepsilon(x)|) < \infty$. Let $0 < \alpha_1 < \alpha_2 < 1$. Then*

$$\left\{ \frac{n^{\beta-1/2}}{\sigma\ell(n)} (\hat{\theta}_\alpha - \theta_\alpha), \alpha_1 \leq \alpha \leq \alpha_2 \right\} \Rightarrow \{c_\beta Z, \alpha_1 \leq \alpha \leq \alpha_2\}, \quad (2.23)$$

where $Z \sim N(0, 1)$ and $\sigma^2 = \mathbb{E}|\varepsilon_i|^2$.

The limiting process in (2.23) is a multiple of a normal random variable and is therefore degenerate. It reveals the very surprising and interesting fact that $\hat{\theta}_\alpha - \theta_\alpha$ has the same asymptotic distribution for all $\alpha \in [\alpha_1, \alpha_2]$. In comparison, in the SRD case, the limiting distribution (2.16) is a non-degenerate Gaussian process.

2.2.5 Application to $G/G/1$ Queues

Dingeç et al. [2022] considered the waiting-time process arising from a steady-state $G/G/1$ queueing system with traffic intensity less than unity and a sub-exponential service-time distribution. Let T_k be the inter-arrival time between the k th and $(k+1)$ th customer and S_k be the service time of the k th customer. Then the waiting time W_k for the k th customer satisfies the Lindley recursion

$$W_{k+1} = G_{\varepsilon_k}(W_k), \text{ where } \varepsilon_k = S_k - T_k \text{ and } G_{\varepsilon}(w) = \max(0, w + \varepsilon). \quad (2.24)$$

Here $G_{\varepsilon_k}(\cdot)$ is a random map. Further assume that $\{T_k\}$ and $\{S_k\}$ are each i.i.d. sequences and they are mutually independent. Then ε_k are also i.i.d. Assume that the moment generating function $\varphi(c) = \mathbb{E}(e^{c\varepsilon_k}) < \infty$ for some $c > 0$ and $\mathbb{E}\varepsilon_k < 0$. Then as $k \rightarrow \infty$, W_k converges to the steady state distribution W_{∞} , which is distributed as

$$X_k = s(\varepsilon_k, \varepsilon_{k-1}, \varepsilon_{k-2}, \dots) = G_{\varepsilon_k} \circ G_{\varepsilon_{k-1}} \circ G_{\varepsilon_{k-2}} \circ \dots, \quad (2.25)$$

where s is a measurable function and $\circ$ stands for function composition. Dineç et al. [2022] proved the GMC property that, for any $\psi > 0$, there exists constants $\rho \in (0, 1)$ and $C > 0$ such that for all $k \geq 0$,

$$\|X_k - X'_k\|_{\psi} \leq C\rho^k. \quad (2.26)$$

The above property implies that the functional dependence measure decays to 0 geometrically quickly, and hence (2.9) and (2.14) are satisfied. Thus our Theorems 2.2.1 and 2.2.2 are applicable and we can have the asymptotic normality and weak convergence of expectile estimates for steady state distributions.

2.3 Asymptotic Theory for Expected Shortfall

Remind that $(X_t)_{t \in \mathbb{Z}}$ is a stationary casual process and the shift process $\mathcal{F}_t = \{\dots, \varepsilon_{t-1}, \varepsilon_t\}$, where ε_k , $k \in \mathbb{Z}$, are independent and identically distributed (i.i.d.) random variables. Now we denote the marginal distribution and density function of X_t by F and f respectively. For $i \in \mathbb{N}$, let $F_i(u|\mathcal{F}_j) = \mathbb{P}(X_{i+j} \leq u|\mathcal{F}_j)$ [resp., $f_i(u|\mathcal{F}_j)$] be the conditional distribution [resp., density] function of X_{i+j} at u given $\mathcal{F}_j$. Remind that $(\varepsilon'_i)_{i \in \mathbb{Z}}$ is an i.i.d. copy of $(\varepsilon_i)_{i \in \mathbb{Z}}$, $\mathcal{F}_t^* = \{\dots, \varepsilon_{-1}, \varepsilon'_0, \varepsilon_1, \dots, \varepsilon_t\}$ and $X_t^* = s(\mathcal{F}_t^*)$.

For $0 \leq \alpha \leq 1$, denote by $q_\alpha = \inf\{x : F(x) \geq \alpha\}$ the α -th quantile of F , and $\text{ES}_\alpha = -\mathbb{E}[X_1 \mathbb{1}_{X_1 \leq q_\alpha}]/\alpha$ the α -th expected shortfall. Given a sample $X_1, \dots, X_n$, let $\hat{q}_\alpha$ be the α -th sample quantile

$$\hat{q}_\alpha = \arg \min_{\beta} \sum_{i=1}^n \rho_\alpha(X_i - \beta),$$

where $\rho_\alpha(u) = \alpha u^+ + (1 - \alpha)(-u)^+$ and $u^+ = \max(0, u)$. We also refer to $\hat{q}_\alpha$ as the α -th quantile of the empirical distribution function $F_n(x) = \sum_{i=1}^n \mathbb{1}_{X_i \leq x}/n$. Let $\widehat{\text{ES}}_\alpha$ be the α -th sample expected shortfall

$$\widehat{\text{ES}}_\alpha = -\frac{1}{\alpha n} \sum_{i=1}^n X_i \mathbb{1}_{X_i \leq \hat{q}_\alpha}.$$

2.3.1 ES for short range dependent processes

Before we study ES, it's meaningful to present some results on quantiles. Under SRD conditions, Wu [2007b] developed powerful theories about quantiles, including strong Bahadur representation.

Lemma 2.3.1. *Assume that $f(q_\alpha) > 0$, $f \in \mathcal{C}^1$ and there is a constant C such that $\sup_u f_1(u|\mathcal{F}_0) \leq C < \infty$, $\sup_u \|F_1^{(l)}(u|\mathcal{F}_i)\| < \infty$ and either (i)*

$$\sum_{i=0}^{\infty} \sup_{u \in \mathbb{R}} \|\mathbb{E}[F_1^{(l)}(u|\mathcal{F}_i)|\mathcal{F}_0] - \mathbb{E}[F_1^{(l)}(u|\mathcal{F}_i^*)|\mathcal{F}_0^*]\| < \infty \quad (2.27)$$

or (ii)

$$\sum_{i=0}^{\infty} \sup_{u \in \mathbb{R}} \|F_1^{(l)}(u|\mathcal{F}_i) - F_1^{(l)}(u|\mathcal{F}_i^*)\| < \infty \quad (2.28)$$

holds for $l = 0, 1$. Then the strong Bahadur representation holds:

$$\hat{q}_\alpha - q_\alpha - \frac{\alpha - F_n(q_\alpha)}{f(q_\alpha)} = O_{\text{a.s.}}(n^{-3/4}(\log n)^{q'}) \quad (2.29)$$

for some $q' > 9/4$.

Lemma 2.3.2. *With the assumptions in Lemma 2.3.1, let $b_n = n^{-1/2}(\log n)^{3/2}(\log \log n)^{1/2+\iota}$, $\iota > 0$. Then we have (i)*

$$\sup_{|x - q_\alpha| \leq b_n} |[F_n(x) - F(x)] - [F_n(q_\alpha) - F(q_\alpha)]| = O_{\text{a.s.}}(n^{-3/4}(\log n)^{q'}) \quad (2.30)$$

and (ii)

$$|F_n(q_\alpha) - \alpha| = o_{\text{a.s.}}(n^{-1/2}(\log n)^{3/2}(\log \log n)^{1/2+\iota/4}) \quad (2.31)$$

Lemma 2.3.3. *Let $H_n(\theta) = \sum_{i=1}^n \{X_i \mathbb{1}_{X_i \leq \theta} - \mathbb{E}[X_i \mathbb{1}_{X_i \leq \theta}]\}$. With the assumptions of Lemma 2.3.1 and b_n in Lemma 2.3.2, for any $\theta_0 \in \mathbb{R}$, we have*

$$\sup_{|\theta - \theta_0| \leq b_n} |H_n(\theta) - H_n(\theta_0)| = O_{\text{a.s.}}(|\theta_0| n^{1/4}(\log n)^{q'}) \quad (2.32)$$

Now we can present the central limit theorem and weak convergence of $\widehat{\text{ES}}_\alpha$. These theorems have typically the same logic as those for expectiles. The first one is the CLT for $\widehat{\text{ES}}_\alpha$ at single level and the second one is the weak convergence.

Theorem 2.3.4. *Let $g(\theta) = \mathbb{E}X_1 \mathbb{1}_{X_1 \leq \theta} \in \mathcal{C}^2$ and $h_\alpha(X_i) = (X_i - q_\alpha) \mathbb{1}_{X_i \leq q_\alpha}$. Under the assumptions of Lemma 2.3.1, 2.3.2, 2.3.3 and $\sup_\theta [|g'(\theta)| + |g''(\theta)|] < \infty$, then, for any fixed $\alpha \in (0, 1)$,*

$$\sqrt{n}\alpha \left\{ \widehat{\text{ES}}_\alpha - \text{ES}_\alpha \right\} \xrightarrow{d} N(0, \sigma_\infty^2) \quad (2.33)$$

where $\sigma_\infty^2 = \sum_{i=-\infty}^{\infty} \text{Cov}(h_\alpha(X_0), h_\alpha(X_i))$.

Theorem 2.3.5. *Suppose that $\inf_{\alpha_1 \leq \alpha \leq \alpha_2} f(q_\alpha) > 0$ for $0 < \alpha_1 < \alpha_2 < 1$. Under the assumptions of Theorem 2.3.4, we have the weak convergence*

$$\left\{ \sqrt{n} \alpha \left[\widehat{ES}_\alpha - ES_\alpha \right], \alpha_1 \leq \alpha \leq \alpha_2 \right\} \xrightarrow{d} \{W(\alpha), \alpha_1 \leq \alpha \leq \alpha_2\}$$

where $W(\alpha)$ is a mean 0 Gaussian process with covariance function

$$\mathbb{E}(W(\alpha)W(\alpha')) = \sum_{i=-\infty}^{\infty} \text{Cov}(h_\alpha(X_0), h_{\alpha'}(X_i)).$$

2.3.2 ES for long-memory processes

We also consider the linear process

$$X_t = \sum_{i=0}^{\infty} a_i \varepsilon_{t-i}, \quad (2.34)$$

where ε_t are i.i.d. with mean 0, $a_0 = 1$, $a_n = n^{-\beta} \ell(n)$, $n \geq 1$, where $1/2 < \beta < 1$, and ℓ is a slowly varying function. Let F_ε and f_ε be the distribution function and density of ε_t . Let $\sigma^2 = \mathbb{E}|\varepsilon_i|^2$. When $\alpha = 1$, i.e. $\widehat{ES}_\alpha = \bar{X}_n$, we have already mentioned that

$$\frac{\bar{X}_n}{n^{1/2-\beta} \ell(n) \sigma} \xrightarrow{d} N(0, c_\beta^2), \text{ where } c_\beta^2 = \int_0^\infty \left(\int_{t-1}^t \max(u, 0)^{-\beta} du \right)^2 dt. \quad (2.35)$$

Let $\Psi_n = \sqrt{n} \sum_{k=1}^n k^{1/2-2\beta} L^2(k)$, and $A_n(\beta) = \Psi_n^2(\log n)(\log \log n)^2$ if $\beta < 3/4$ and $A_n(\beta) = \Psi_n^2(\log n)^3(\log \log n)^2$ if $\beta \geq 3/4$. However, being different from expectile, ES estimators are not always asymptotically normal. For some special α , which depends on the distribution of ε_i , $\widehat{ES}_\alpha$ is asymptotically Rosenblatt distributed.

Theorem 2.3.6. *Assume $\inf_{\alpha_1 \leq \alpha \leq \alpha_2} f(q_\alpha) > 0$ for some $0 < \alpha_1 < \alpha_2 < 1$, $\mathbb{E}(\varepsilon_1)^4 < \infty$*

and

$$\sup_x \{|f_\varepsilon(x)| + |f'_\varepsilon(x)|\} + \int_{\mathbb{R}} |f'_\varepsilon(u)|^2 du < \infty.$$

Assume that $f(x), g(\theta) \in \mathcal{C}^2$, $\int_{-\infty}^{\infty} |xf'(x)|dx < \infty$, $\int_{-\infty}^{\infty} |xf''(x)|dx < \infty$ and $\int_{-\infty}^{\theta} xf''(x)dx \neq 0$ for any $\theta \in \mathbb{R}$. Then

(i) if $\kappa_{1,\alpha} = \int_{-\infty}^{q_\alpha} xf'(x)dx \neq 0$, then

$$\frac{\widehat{ES}_\alpha - ES_\alpha}{n^{1/2-\beta}\ell(n)\sigma} \xrightarrow{d} N(0, \kappa_{1,\alpha}^2). \quad (2.36)$$

(ii) if $\kappa_{1,\alpha} = \int_{-\infty}^{q_\alpha} xf'(x)dx = 0$ and $\kappa_{2,\alpha} = \int_{-\infty}^{q_\alpha} xf''(x)dx \neq 0$, then when $\beta < 3/4$,

$$\frac{\widehat{ES}_\alpha - ES_\alpha}{n^{1/2-\beta}\ell(n)\sigma} \xrightarrow{d} \kappa_{2,\alpha} \int_{-\infty}^1 \int_{u_1}^1 \int_0^1 [(v - u_1)^+ \cdot (v - u_2)^+]^{-\beta} dv dB(u_2) dB(u_1), \quad (2.37)$$

where $x^+ = \max(x, 0)$ and the asymptotic distribution is Rosenblatt distribution. When $\beta > 3/4$ or $\beta = 3/4$ and $\sum_{n \geq 1} |l^2(n)|/n < \infty$, then

$$\sqrt{n}(\widehat{ES}_\alpha - ES_\alpha) \xrightarrow{d} N(0, \sigma^2), \quad (2.38)$$

for some constant $\sigma < \infty$.

2.4 Simulation

2.4.1 Expectile Estimation for SRD Processes

CI Simulation

This section implements CI simulation for AR processes, TAR processes and G/G/1 systems using Corollary 2.2.1.1. Following Schmeiser [1982], Steiger and Wilson [2002] and Alexopoulos and Goldsman [2004], we adopt the batching method, also known as the d -fold

method, for CI simulation. In their setting, batching is used for calculating CIs for the mean based on the sample mean which is a linear form of the sequence arising from the simulation experiment. In our problem, the expectile estimator is a nonlinear function and the specification allows for both linear and nonlinear processes. Nonetheless, due to the Bahadur representation which asserts that the difference between the estimated expectiles and the true expectiles is approximately a linear form when $\varphi(u)$ has a derivative at 0; cf (A.3) and (A.4), we can still apply the batching method to construct CI. This section also numerically validates the application of the batching method to expectiles within this context and provides recommendations for selecting the most appropriate d values corresponding to different sample sizes. In the following simulation results, for all types of processes, the empirical coverage probabilities and average CI half-lengths are computed based on 10,000 independent simulation replications.

AR Processes. We first consider the autoregressive AR(1) time series model

$$X_i = \phi X_{i-1} + \varepsilon_i, i \in \mathbb{Z},$$

where the autoregressive parameter $\phi \in (-1, 1)$, and the error terms ε_i are i.i.d. In our simulation study we let ε_i follow Student's t distribution, $\phi = 0.5$, and, as stated before, the empirical coverage probabilities are computed based on 10,000 replications. We estimate the sample α th expectile based on the realization $X_1, \dots, X_n$.

Table 2.1 shows the empirical coverage probabilities of two-sided 95% expectile CIs for the AR(1) process with $t(3)$ innovations. As expected, the coverage probabilities are closer to the nominal level 0.95 as the sample size increases within a fixed batch d . In line with Schmeiser [1982], $d = 2$ is the best choice in terms of coverage probabilities, especially when the sample size is small. The d -fold method generally performs better when d is smaller, as Goldsman and Schruben [1984] find that the interval estimators may not perform well when

Table 2.1: Empirical coverage probabilities of two-sided 95% CIs of expectiles for the AR(1) process with $t(3)$ innovations using the d -fold method with sample size n and expectile level α

d	n	$\alpha = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	100	0.944	0.947	0.949	0.956	0.956	0.952	0.951	0.951	0.951
	200	0.949	0.952	0.954	0.953	0.951	0.952	0.950	0.951	0.949
	500	0.953	0.950	0.950	0.951	0.949	0.951	0.950	0.954	0.950
	1000	0.948	0.947	0.949	0.949	0.950	0.951	0.952	0.951	0.948
	5000	0.953	0.952	0.946	0.948	0.950	0.949	0.951	0.951	0.954
3	100	0.933	0.941	0.945	0.949	0.951	0.954	0.950	0.948	0.932
	200	0.941	0.949	0.950	0.949	0.950	0.952	0.951	0.948	0.941
	500	0.945	0.949	0.950	0.949	0.953	0.953	0.949	0.947	0.946
	1000	0.948	0.950	0.949	0.950	0.949	0.948	0.949	0.947	0.945
	5000	0.948	0.952	0.953	0.955	0.953	0.951	0.948	0.948	0.948
4	100	0.915	0.936	0.944	0.945	0.946	0.945	0.942	0.935	0.915
	200	0.933	0.945	0.948	0.950	0.951	0.950	0.948	0.943	0.928
	500	0.937	0.946	0.948	0.950	0.949	0.947	0.948	0.946	0.939
	1000	0.944	0.946	0.946	0.948	0.950	0.950	0.948	0.947	0.941
	5000	0.949	0.951	0.952	0.953	0.952	0.951	0.950	0.949	0.948
5	100	0.901	0.930	0.942	0.948	0.949	0.948	0.943	0.932	0.903
	200	0.923	0.944	0.949	0.952	0.951	0.952	0.951	0.945	0.925
	500	0.932	0.944	0.949	0.948	0.952	0.951	0.949	0.944	0.934
	1000	0.941	0.946	0.949	0.947	0.949	0.949	0.948	0.945	0.936
	5000	0.949	0.950	0.949	0.949	0.949	0.949	0.948	0.949	0.945
6	100	0.884	0.923	0.939	0.941	0.944	0.946	0.939	0.926	0.887
	200	0.910	0.936	0.942	0.946	0.950	0.949	0.947	0.938	0.912
	500	0.929	0.940	0.949	0.950	0.949	0.947	0.945	0.940	0.930
	1000	0.936	0.944	0.946	0.949	0.950	0.950	0.949	0.945	0.938
	5000	0.947	0.950	0.951	0.951	0.949	0.950	0.949	0.947	0.943
10	200	0.876	0.919	0.936	0.946	0.949	0.948	0.940	0.925	0.879
	500	0.913	0.935	0.944	0.949	0.950	0.948	0.944	0.937	0.914
	1000	0.926	0.940	0.943	0.945	0.949	0.949	0.945	0.941	0.929
	10000	0.947	0.948	0.948	0.948	0.949	0.948	0.949	0.948	0.947
20	1000	0.906	0.929	0.936	0.941	0.941	0.942	0.940	0.933	0.907
	5000	0.938	0.943	0.944	0.948	0.949	0.948	0.948	0.943	0.937
	10000	0.945	0.947	0.950	0.949	0.950	0.950	0.949	0.949	0.945
	100000	0.949	0.950	0.949	0.947	0.947	0.947	0.948	0.946	0.948
100	5000	0.888	0.919	0.934	0.943	0.945	0.945	0.936	0.918	0.886
	10000	0.923	0.936	0.944	0.947	0.949	0.947	0.945	0.939	0.928
	100000	0.946	0.946	0.947	0.946	0.948	0.948	0.945	0.947	0.948

Table 2.2: Halfwidths (multiplied by $\sqrt{n}$) of two-sided 95% CIs of expectiles for the AR(1) process with innovations $t(3)$ using the d -fold method with sample size n and expectile level α

d	n	$\alpha = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	100	50.437	39.748	35.838	34.206	33.803	34.421	36.358	40.712	52.150
	200	55.052	41.818	37.051	34.950	34.246	34.707	36.575	41.085	53.903
	500	56.504	42.358	37.411	35.324	34.749	35.382	37.458	42.307	56.289
	1000	57.164	42.591	37.674	35.620	35.040	35.632	37.750	42.723	57.129
	5000	58.697	43.592	38.383	36.144	35.459	36.025	38.148	43.218	58.054
3	100	18.372	14.727	13.312	12.693	12.531	12.747	13.422	14.925	18.760
	200	20.044	15.417	13.722	12.990	12.777	12.987	13.712	15.409	20.076
	500	20.987	15.703	13.870	13.099	12.879	13.097	13.854	15.644	20.815
	1000	21.524	15.992	14.091	13.284	13.043	13.262	14.045	15.892	21.317
	5000	21.737	16.153	14.246	13.434	13.193	13.413	14.210	16.113	21.704
4	100	13.666	11.109	10.094	9.648	9.529	9.685	10.178	11.259	13.940
	200	15.140	11.739	10.476	9.928	9.769	9.927	10.476	11.743	15.142
	500	16.142	12.097	10.686	10.084	9.906	10.069	10.652	12.029	15.983
	1000	16.428	12.225	10.800	10.206	10.039	10.214	10.814	12.235	16.412
	5000	16.741	12.423	10.958	10.333	10.145	10.310	10.916	12.367	16.662
5	100	11.818	9.737	8.884	8.502	8.400	8.533	8.951	9.855	12.035
	200	13.324	10.431	9.327	8.840	8.697	8.837	9.321	10.427	13.336
	500	14.321	10.756	9.498	8.961	8.801	8.945	9.462	10.687	14.177
	1000	14.605	10.858	9.591	9.063	8.917	9.078	9.621	10.899	14.640
	5000	14.882	11.019	9.708	9.152	8.988	9.139	9.683	10.983	14.826
6	100	10.683	8.889	8.132	7.792	7.702	7.823	8.196	8.997	10.878
	200	12.312	9.696	8.681	8.232	8.100	8.228	8.674	9.688	12.308
	500	13.409	10.098	8.921	8.417	8.268	8.403	8.889	10.036	13.277
	1000	13.714	10.174	8.966	8.459	8.313	8.457	8.961	10.162	13.675
	5000	14.000	10.349	9.114	8.591	8.437	8.579	9.091	10.312	13.930
10	200	10.423	8.456	7.647	7.280	7.173	7.283	7.654	8.469	10.446
	500	11.794	8.986	7.951	7.504	7.373	7.499	7.938	8.958	11.732
	1000	12.364	9.176	8.071	7.605	7.470	7.599	8.059	9.152	12.311
	10000	12.652	9.330	8.212	7.741	7.605	7.738	8.207	9.326	12.648
20	1000	11.373	8.551	7.522	7.082	6.955	7.080	7.517	8.539	11.341
	5000	12.029	8.765	7.683	7.232	7.103	7.229	7.676	8.752	12.001
	10000	12.005	8.781	7.710	7.263	7.135	7.262	7.709	8.780	12.010
	100000	11.935	8.792	7.728	7.279	7.150	7.276	7.723	8.785	11.933
100	5000	11.396	8.384	7.299	6.839	6.709	6.842	7.305	8.395	11.418
	10000	11.817	8.482	7.363	6.900	6.769	6.902	7.367	8.492	11.842
	100000	11.659	8.456	7.403	6.965	6.839	6.963	7.401	8.452	11.657

the number of observations in each fold is small since the Gaussian approximation can be bad due to a relatively small batch size.

Table 2.2 shows the halfwidths for the same process. The halfwidths are multiplied by $\sqrt{n}$ to adjust for the impact of sample size. As Schmeiser [1982] illustrated, the more batches used, the less information lost and the shorter expected halfwidth. On the other hand, the Gaussian approximation can be worse due to small batch size. Table 2.2 numerically reveals this trend where the halfwidths decrease as d increases. Notably, we can observe a substantial reduction in halfwidth when d increases from 2 to 3. We will discuss the tradeoff between the halfwidth and coverage in detail at the end of this subsection, which is crucial for selecting the number of batches d .

TAR Processes. Consider the first-order TAR model

$$X_i = \phi|X_{i-1}| + \varepsilon_i, \quad (2.39)$$

where ε_i are i.i.d. and $-1 < \phi < 1$. In our specification $\phi = -0.5$. Anděl et al. [1984] provide the marginal probability density $h(x)$ for (2.39), where ε_i are independent $\mathbf{N}(0, \sigma^2)$ and $-1 < \phi < 0$,

$$h(x) = [2(1 - \phi^2)/\pi]^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(1 - \phi^2)x^2/\sigma^2\right\} \Phi(-\phi x/\sigma),$$

with $\Phi(\cdot)$ the cumulative distribution function (c.d.f.) of the standard normal distribution.

We implement CI simulation for the TAR model with $\mathbf{N}(0, 1)$ and $t(3)$ innovations, and investigate the finite-sample performance by considering different choices of the level α , the sample size n and the number of batches d . Tables 2.3 and 2.5 display empirical coverage probabilities for the TAR(1) process with $\mathbf{N}(0, 1)$ and $t(3)$ innovations, respectively, while Tables 2.4 and 2.6 present the corresponding scaled halfwidths. The coverage probabilities

Table 2.3: Empirical coverage probabilities of two-sided 95% CIs of expectiles for the TAR(1) process with innovations $\mathbf{N}(0, 1)$ using the d -fold method with sample size n and expectile level α

d	n	$\alpha = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	100	0.950	0.950	0.951	0.950	0.954	0.953	0.950	0.950	0.953
	200	0.949	0.950	0.951	0.949	0.950	0.952	0.953	0.949	0.949
	500	0.948	0.952	0.950	0.951	0.950	0.948	0.947	0.951	0.945
	1000	0.952	0.951	0.953	0.955	0.950	0.950	0.950	0.949	0.949
	5000	0.949	0.950	0.948	0.951	0.950	0.949	0.950	0.950	0.954
3	100	0.946	0.947	0.948	0.951	0.950	0.950	0.951	0.951	0.951
	200	0.955	0.953	0.952	0.950	0.951	0.952	0.955	0.952	0.950
	500	0.950	0.948	0.946	0.946	0.946	0.947	0.946	0.947	0.947
	1000	0.952	0.951	0.952	0.949	0.950	0.951	0.948	0.949	0.950
	5000	0.949	0.948	0.948	0.946	0.949	0.948	0.947	0.948	0.950
4	100	0.940	0.945	0.948	0.949	0.948	0.948	0.948	0.950	0.952
	200	0.945	0.950	0.950	0.949	0.950	0.951	0.950	0.952	0.952
	500	0.944	0.942	0.943	0.944	0.944	0.945	0.948	0.948	0.948
	1000	0.950	0.948	0.948	0.948	0.948	0.948	0.948	0.950	0.948
	5000	0.952	0.954	0.954	0.954	0.955	0.954	0.955	0.953	0.952
5	100	0.935	0.942	0.946	0.945	0.948	0.947	0.949	0.952	0.952
	200	0.941	0.943	0.944	0.944	0.945	0.948	0.950	0.954	0.953
	500	0.944	0.945	0.944	0.944	0.945	0.946	0.947	0.946	0.946
	1000	0.950	0.947	0.948	0.949	0.949	0.949	0.951	0.951	0.947
	5000	0.950	0.950	0.948	0.949	0.950	0.951	0.952	0.951	0.948
6	100	0.926	0.937	0.942	0.945	0.949	0.950	0.950	0.952	0.951
	200	0.938	0.945	0.949	0.948	0.951	0.951	0.952	0.954	0.952
	500	0.946	0.948	0.947	0.948	0.947	0.948	0.945	0.944	0.945
	1000	0.948	0.947	0.948	0.948	0.949	0.949	0.950	0.950	0.949
	5000	0.949	0.949	0.951	0.951	0.951	0.952	0.952	0.951	0.950
10	200	0.918	0.932	0.940	0.947	0.951	0.950	0.952	0.951	0.947
	1000	0.943	0.946	0.947	0.947	0.947	0.948	0.946	0.946	0.946
	5000	0.949	0.950	0.951	0.951	0.952	0.951	0.952	0.950	0.949
	10000	0.949	0.949	0.949	0.948	0.948	0.947	0.949	0.950	0.948
20	1000	0.924	0.936	0.941	0.944	0.946	0.946	0.945	0.944	0.940
	5000	0.949	0.950	0.949	0.950	0.949	0.950	0.947	0.945	0.946
	10000	0.953	0.953	0.953	0.952	0.949	0.949	0.948	0.951	0.951
	100000	0.953	0.952	0.951	0.953	0.954	0.951	0.952	0.950	0.950
100	5000	0.855	0.914	0.936	0.945	0.946	0.945	0.941	0.929	0.909
	10000	0.900	0.929	0.940	0.945	0.948	0.947	0.945	0.941	0.929
	100000	0.948	0.951	0.949	0.951	0.951	0.951	0.949	0.948	0.947

Table 2.4: Halfwidths (multiplied by $\sqrt{n}$) of two-sided 95% CIs of expectiles for the TAR(1) process with innovations $\mathbf{N}(0, 1)$ using the d -fold method with sample size n and expectile level α

d	n	$\alpha = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	100	17.222	15.221	14.197	13.523	13.031	12.666	12.441	12.447	13.174
	200	17.071	15.058	14.022	13.333	12.825	12.452	12.205	12.170	12.903
	500	17.304	15.172	14.082	13.358	12.818	12.407	12.128	12.077	12.788
	1000	17.518	15.408	14.320	13.598	13.054	12.634	12.328	12.252	12.937
	5000	17.457	15.317	14.225	13.503	12.959	12.549	12.270	12.224	12.970
3	100	6.380	5.663	5.291	5.045	4.866	4.736	4.659	4.667	4.949
	200	6.465	5.715	5.319	5.054	4.858	4.709	4.608	4.588	4.840
	500	6.451	5.682	5.290	5.031	4.841	4.697	4.599	4.582	4.843
	1000	6.518	5.739	5.339	5.073	4.876	4.726	4.622	4.605	4.870
	5000	6.626	5.819	5.407	5.132	4.928	4.770	4.654	4.621	4.872
4	100	4.823	4.298	4.022	3.839	3.707	3.611	3.555	3.568	3.789
	200	4.916	4.355	4.064	3.871	3.731	3.626	3.558	3.552	3.759
	500	4.978	4.381	4.075	3.872	3.724	3.612	3.538	3.530	3.740
	1000	5.022	4.420	4.110	3.902	3.747	3.627	3.543	3.524	3.719
	5000	5.062	4.453	4.140	3.931	3.775	3.653	3.566	3.544	3.744
5	100	4.249	3.804	3.567	3.410	3.297	3.218	3.175	3.195	3.407
	200	4.353	3.860	3.602	3.431	3.306	3.213	3.152	3.150	3.338
	500	4.431	3.898	3.624	3.443	3.310	3.209	3.142	3.131	3.314
	1000	4.464	3.926	3.650	3.465	3.328	3.222	3.148	3.132	3.307
	5000	4.500	3.957	3.679	3.493	3.353	3.245	3.167	3.143	3.316
6	100	3.901	3.502	3.292	3.154	3.057	2.990	2.956	2.980	3.179
	200	4.078	3.621	3.382	3.225	3.111	3.026	2.973	2.974	3.149
	500	4.142	3.652	3.399	3.231	3.107	3.014	2.951	2.941	3.110
	1000	4.159	3.666	3.411	3.242	3.117	3.022	2.956	2.944	3.111
	5000	4.209	3.701	3.442	3.269	3.140	3.040	2.968	2.947	3.108
10	200	3.574	3.196	2.996	2.865	2.771	2.705	2.668	2.683	2.856
	1000	3.741	3.297	3.067	2.915	2.803	2.718	2.660	2.650	2.805
	5000	3.782	3.327	3.092	2.936	2.818	2.727	2.662	2.645	2.793
	10000	3.771	3.314	3.081	2.926	2.810	2.720	2.655	2.638	2.788
20	1000	3.471	3.072	2.867	2.731	2.631	2.557	2.508	2.503	2.648
	5000	3.538	3.113	2.896	2.751	2.643	2.559	2.500	2.487	2.627
	10000	3.539	3.113	2.895	2.750	2.641	2.557	2.498	2.484	2.626
	100000	3.546	3.117	2.898	2.752	2.643	2.560	2.500	2.485	2.625
100	5000	3.325	2.941	2.744	2.613	2.517	2.445	2.398	2.394	2.534
	10000	3.362	2.964	2.760	2.624	2.524	2.448	2.395	2.386	2.523
	100000	3.394	2.984	2.774	2.634	2.530	2.450	2.392	2.378	2.514

for the process with $\mathbf{N}(0, 1)$ innovations are quite close to the nominal level 0.95. This outcome is attributed to the favorable characteristics of light-tailed distributions. For the $t(3)$ innovations, the patterns observed are analogous to those in the AR(1) process.

Table 2.5 and Table 2.6 respectively demonstrate the empirical coverage probabilities and halfwidths for the TAR(1) process with $t(3)$ error terms. Similar results can be found here as shown in Table 2.1 and Table 2.2. There exists a trade-off between coverage and width. For example, $d = 2$ brings satisfactory performance in terms of coverage probabilities but is also associated with large widths implying the loss of information. In general, more batches lead to less information loss and less satisfactory coverage probabilities.

We also provide empirical coverage probabilities and scaled halfwidths of two-sided 99% expectile CIs for AR(1) and TAR(1) processes in the additional simulation study detailed in the appendix. Similar results can be found therein.

G/G/1 Queue-Waiting-Time Process This section conducts CI simulation of expectiles on the G/G/1 queueing system following Alexopoulos et al. [2019] and Dengeç et al. [2022]. Theorems 2.2.1 and 2.2.2 provide the theoretical background for estimates of expectiles of steady state distributions, as clarified in Section 2.2.5.

We first discuss the M/M/1 queueing system with the discipline first in first out (FIFO), the arrival rate $\lambda = 0.9$, the service rate $\omega = 1$, and X_i the time spent in the queue by the i th arriving customer before entering service. In this system, the steady-state server utilization is $\rho = \lambda/\omega = 0.9$, and the steady-state c.d.f of X_i

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - \rho & \text{if } x = 0 \\ 1 - \rho e^{-\omega(1-\rho)x} & \text{if } x > 0. \end{cases}$$

Table A.1 displays that the coverage probabilities are close to the nominal level 95% for the

Table 2.5: Empirical coverage probabilities of two-sided 95% CIs of expectiles for the TAR(1) process with innovations $t(3)$ using the d -fold method with sample size n and expectile level α

d	n	$\alpha = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	100	0.943	0.944	0.943	0.947	0.948	0.952	0.954	0.952	0.953
	200	0.943	0.951	0.951	0.951	0.953	0.953	0.954	0.953	0.950
	500	0.945	0.948	0.952	0.953	0.954	0.954	0.952	0.952	0.952
	1000	0.951	0.950	0.948	0.951	0.951	0.953	0.950	0.949	0.951
	5000	0.950	0.951	0.952	0.951	0.950	0.947	0.951	0.952	0.954
3	100	0.920	0.931	0.933	0.941	0.944	0.945	0.949	0.952	0.948
	200	0.934	0.941	0.948	0.950	0.948	0.951	0.950	0.953	0.952
	500	0.940	0.943	0.945	0.949	0.952	0.952	0.951	0.952	0.946
	1000	0.945	0.946	0.947	0.949	0.950	0.951	0.948	0.949	0.947
	5000	0.947	0.948	0.949	0.951	0.949	0.950	0.953	0.949	0.950
4	100	0.896	0.918	0.929	0.936	0.940	0.946	0.949	0.950	0.940
	200	0.919	0.934	0.942	0.944	0.947	0.951	0.951	0.953	0.947
	500	0.932	0.939	0.944	0.945	0.946	0.950	0.950	0.951	0.946
	1000	0.936	0.942	0.941	0.944	0.947	0.947	0.950	0.951	0.946
	5000	0.944	0.947	0.948	0.949	0.949	0.951	0.953	0.950	0.947
5	100	0.877	0.908	0.923	0.932	0.939	0.945	0.953	0.953	0.937
	200	0.906	0.928	0.938	0.944	0.950	0.952	0.954	0.955	0.945
	500	0.924	0.937	0.942	0.948	0.950	0.950	0.953	0.952	0.947
	1000	0.934	0.942	0.944	0.947	0.947	0.948	0.948	0.953	0.946
	5000	0.946	0.947	0.947	0.949	0.950	0.950	0.950	0.950	0.949
6	100	0.861	0.900	0.917	0.930	0.937	0.945	0.951	0.954	0.935
	200	0.894	0.921	0.933	0.941	0.945	0.948	0.951	0.952	0.941
	500	0.921	0.931	0.937	0.944	0.948	0.950	0.951	0.951	0.945
	1000	0.928	0.937	0.939	0.941	0.944	0.948	0.950	0.951	0.946
	5000	0.946	0.946	0.949	0.950	0.950	0.951	0.951	0.951	0.948
10	200	0.856	0.899	0.919	0.931	0.942	0.948	0.953	0.955	0.934
	1000	0.922	0.933	0.938	0.940	0.943	0.944	0.947	0.950	0.944
	5000	0.942	0.945	0.947	0.948	0.950	0.952	0.952	0.953	0.946
	10000	0.944	0.949	0.949	0.949	0.951	0.953	0.953	0.951	0.950
20	1000	0.901	0.925	0.932	0.937	0.941	0.943	0.945	0.946	0.942
	5000	0.935	0.943	0.947	0.946	0.945	0.949	0.949	0.949	0.947
	10000	0.942	0.946	0.949	0.951	0.951	0.951	0.950	0.950	0.949
	100000	0.949	0.949	0.950	0.949	0.950	0.949	0.948	0.947	0.948
100	5000	0.899	0.926	0.934	0.941	0.945	0.946	0.948	0.951	0.952
	10000	0.928	0.938	0.941	0.946	0.948	0.949	0.950	0.950	0.953
	100000	0.946	0.946	0.947	0.948	0.948	0.946	0.946	0.947	0.947

Table 2.6: Halfwidths (multiplied by $\sqrt{n}$) of two-sided 95% CIs of expectiles for the TAR(1) process with innovations $t(3)$ using the d -fold method with sample size n and expectile level α

d	n	$\alpha = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	100	45.697	32.459	26.675	23.443	21.508	20.555	20.729	22.967	32.252
	200	50.377	34.676	28.116	24.425	22.180	20.969	20.849	22.805	32.063
	500	52.229	35.301	28.495	24.711	22.384	21.092	20.951	22.987	32.323
	1000	53.013	35.601	28.703	24.903	22.571	21.260	21.087	23.124	32.671
	5000	54.457	36.322	29.089	25.086	22.622	21.249	21.112	23.172	32.821
3	100	16.440	11.938	9.918	8.759	8.059	7.709	7.777	8.637	12.060
	200	18.295	12.704	10.343	9.023	8.223	7.799	7.800	8.606	12.168
	500	19.426	13.124	10.578	9.169	8.308	7.836	7.790	8.540	12.054
	1000	20.061	13.449	10.811	9.348	8.449	7.941	7.863	8.601	12.148
	5000	20.439	13.655	10.957	9.461	8.539	8.024	7.942	8.668	12.241
4	100	12.242	9.021	7.526	6.657	6.134	5.887	5.961	6.630	9.193
	200	13.847	9.712	7.922	6.911	6.298	5.975	5.980	6.601	9.331
	500	14.926	10.101	8.143	7.056	6.393	6.031	6.001	6.596	9.341
	1000	15.296	10.254	8.246	7.139	6.466	6.094	6.055	6.637	9.373
	5000	15.646	10.446	8.387	7.245	6.545	6.153	6.097	6.664	9.420
5	100	10.542	7.896	6.636	5.895	5.450	5.241	5.317	5.928	8.165
	200	12.132	8.604	7.049	6.163	5.624	5.345	5.360	5.939	8.399
	500	13.254	8.990	7.244	6.272	5.681	5.360	5.337	5.871	8.359
	1000	13.593	9.090	7.304	6.324	5.731	5.407	5.380	5.913	8.399
	5000	14.011	9.333	7.484	6.459	5.829	5.474	5.419	5.927	8.399
6	100	9.521	7.220	6.095	5.430	5.028	4.845	4.927	5.504	7.542
	200	11.174	7.983	6.555	5.737	5.241	4.987	5.012	5.572	7.895
	500	12.399	8.440	6.808	5.899	5.347	5.046	5.027	5.536	7.892
	1000	12.749	8.530	6.850	5.925	5.363	5.054	5.023	5.519	7.874
	5000	13.107	8.720	6.993	6.039	5.456	5.130	5.085	5.570	7.906
10	200	9.414	6.953	5.784	5.098	4.681	4.476	4.527	5.063	7.089
	1000	11.502	7.707	6.179	5.338	4.829	4.551	4.530	4.997	7.207
	5000	11.856	7.846	6.280	5.419	4.893	4.600	4.562	5.006	7.147
	10000	11.918	7.904	6.327	5.456	4.923	4.622	4.575	5.007	7.116
20	1000	10.559	7.197	5.779	4.991	4.515	4.259	4.255	4.740	6.951
	5000	11.301	7.414	5.909	5.088	4.591	4.315	4.280	4.707	6.783
	10000	11.298	7.428	5.928	5.107	4.607	4.328	4.291	4.711	6.746
	100000	11.292	7.475	5.971	5.140	4.629	4.340	4.294	4.702	6.691
100	5000	10.710	7.148	5.668	4.856	4.369	4.108	4.103	4.612	6.996
	10000	11.169	7.240	5.710	4.886	4.393	4.123	4.100	4.565	6.847
	100000	11.069	7.206	5.728	4.921	4.429	4.152	4.110	4.511	6.482

M/M/1 sequence in almost all cases. The halfwidths decrease with a larger batch d as shown in Table A.2 in line with the previous results.

The second test process is the sequence $\{X_i\}$ of queue-waiting times in an M/H₂/1 queueing system with the queue discipline FIFO, empty-and-idle initial system condition, and the arrival rate $\lambda = 1.0$. The service times have a hyperexponential distributions with mixing probabilities $g = (5 + \sqrt{15})/10 \approx 0.8873$ and $1 - g$ and with the respective service rates $\omega_1 = 2g\alpha$ and $\omega_2 = 2(1 - g)\alpha$, where $\alpha = 1.25$. Accordingly, the steady-state server utilization is $\rho = 0.8$, and the service times have a mean of 0.8 and a coefficient of variance of 2. Results for the M/H₂/1 process are presented in Table A.3 and Table A.4. The coverage probabilities with different combinations of batches, expectile levels and sample sizes are approximately 95%. The halfwidths are scaled by $\sqrt{n}$ to consider the influence of the sample size, the tendency of which in Table A.4 is similar to that in Table A.2.

The third test process is the sequence $\{X_i\}$ of queue-waiting times in an M/G/1 queueing system with service time S following a lognormal distribution with parameters μ and σ , i.e., $\log S \sim \mathbf{N}(\mu, \sigma^2)$. The steady-state density of X_i is $f(x) = e^{-\mu - \sigma^2/2} \Phi[-(\log x - \mu)/\sigma]$. Table A.5 reports the best coverage probabilities with $d = 2$, while the half-widths are much larger with $d = 2$ as shown in Table A.6. There is a trade-off consistent with the previous results.

With the support of our theorems, batching can be employed for computing CIs on the expectiles by transforming correlated observations into less uncorrelated and approximately normally distributed batch means. To provide more information for selecting a satisfactory d , we also calculate grand expectiles in Table 2.7 alongside the true expectiles and relative precision (%) in Table 2.8 using different values of α compared to the previous tables. Relative precision can be regarded as the margin of error as clarified in Alexopoulos et al. [2019]. A larger d is usually associated with smaller relative precision with other conditions being equal, and it can lead to a smaller standard deviation for the half-length, as we

Table 2.7: Grand expectiles of two-sided 95% expectile CIs using the d -fold method with sample size n and expectile level α

AR(1) process with innovations $t(3)$								
d	n	$\alpha = 0.05$	0.25	0.45	0.65	0.85	0.95	0.99
		-2.198	-0.751	-0.135	0.420	1.211	2.197	4.017
6	10000	-2.199	-0.751	-0.135	0.420	1.211	2.200	4.047
	100000	-2.197	-0.751	-0.135	0.420	1.211	2.198	4.024
20	10000	-2.202	-0.750	-0.135	0.419	1.210	2.202	4.051
	100000	-2.198	-0.751	-0.135	0.420	1.211	2.198	4.032
100	10000	-2.182	-0.745	-0.134	0.416	1.202	2.182	3.820
	100000	-2.200	-0.751	-0.135	0.419	1.211	2.201	4.055
TAR(1) process with innovations $\mathbf{N}(0, 1)$								
d	n	$\alpha = 0.05$	0.25	0.45	0.65	0.85	0.95	0.99
		-1.676	-0.924	-0.545	-0.200	0.261	0.739	1.341
6	10000	-1.675	-0.923	-0.545	-0.200	0.261	0.738	1.340
	100000	-1.676	-0.924	-0.545	-0.200	0.261	0.739	1.341
20	10000	-1.673	-0.923	-0.545	-0.200	0.260	0.737	1.337
	100000	-1.676	-0.924	-0.545	-0.200	0.261	0.739	1.340
100	10000	-1.660	-0.919	-0.545	-0.202	0.257	0.731	1.321
	100000	-1.674	-0.923	-0.545	-0.200	0.261	0.738	1.339
M/M/1 process								
d	n	$\alpha = 0.05$	0.25	0.45	0.65	0.85	0.95	0.99
		2.063	5.558	8.285	11.453	16.815	23.877	35.178
6	10000	2.065	5.559	8.285	11.452	16.811	23.868	35.141
	100000	2.063	5.558	8.285	11.454	16.815	23.877	35.173
20	10000	2.068	5.561	8.285	11.450	16.803	23.846	35.069
	100000	2.064	5.559	8.285	11.453	16.814	23.875	35.165
100	10000	2.089	5.573	8.288	11.439	16.757	23.719	34.595
	100000	2.066	5.560	8.286	11.452	16.810	23.862	35.124

Table 2.8: Relative precision (%) of two-sided 95% expectile CIs using the d -fold method with sample size n and expectile level α

AR(1) process with innovations $t(3)$								
d	n	$\alpha = 0.05$	0.25	0.45	0.65	0.85	0.95	0.99
6	10000	9.335	12.794	62.816	21.005	9.592	9.380	15.708
	100000	2.987	4.067	19.894	6.640	3.037	2.973	5.010
20	10000	8.163	10.822	52.986	17.731	8.163	8.196	13.422
	100000	2.540	3.429	16.789	5.613	2.575	2.540	4.435
100	10000	7.985	10.469	50.714	17.024	8.021	8.007	10.328
	100000	2.554	3.292	16.072	5.378	2.490	2.554	4.672
TAR(1) process with innovations $\mathbf{N}(0, 1)$								
d	n	$\alpha = 0.05$	0.25	0.45	0.65	0.85	0.95	0.99
6	10000	2.889	3.846	5.853	14.940	11.429	4.759	4.025
	100000	0.909	1.214	1.848	4.724	3.603	1.494	1.263
20	10000	2.431	3.242	4.935	12.583	9.660	4.020	3.400
	100000	0.768	1.025	1.562	3.995	3.051	1.268	1.071
100	10000	2.318	3.099	4.720	11.980	9.428	3.892	3.278
	100000	0.736	0.982	1.496	3.820	2.925	1.217	1.029
M/M/1 process								
d	n	$\alpha = 0.05$	0.25	0.45	0.65	0.85	0.95	0.99
6	10000	3.988	2.897	2.697	2.659	2.806	3.239	4.411
	100000	1.276	0.930	0.864	0.849	0.894	1.030	1.400
20	10000	3.393	2.470	2.299	2.266	2.392	2.758	3.754
	100000	1.075	0.782	0.727	0.716	0.755	0.870	1.182
100	10000	3.267	2.371	2.207	2.175	2.295	2.646	3.598
	100000	1.030	0.749	0.697	0.687	0.724	0.834	1.135

Table 2.9: Empirical coverage probabilities of two-sided 95% expectile CIs using the d -fold method with sample size n and expectile level α

d	n	$\alpha = 0.05$	0.25	0.45	0.65	0.85	0.95	0.99
AR(1) process with innovations $t(3)$								
6	10000	0.945	0.951	0.950	0.948	0.948	0.946	0.928
	100000	0.946	0.948	0.948	0.948	0.947	0.949	0.944
20	10000	0.938	0.949	0.949	0.950	0.949	0.942	0.917
	100000	0.944	0.950	0.947	0.947	0.946	0.950	0.947
100	10000	0.906	0.940	0.948	0.947	0.935	0.914	0.707
	100000	0.946	0.946	0.948	0.946	0.947	0.950	0.953
TAR(1) process with innovations $\mathbf{N}(0, 1)$								
d	n	$\alpha = 0.05$	0.25	0.45	0.65	0.85	0.95	0.99
6	10000	0.953	0.947	0.945	0.946	0.948	0.949	0.950
	100000	0.953	0.952	0.950	0.950	0.949	0.949	0.951
20	10000	0.946	0.948	0.948	0.947	0.948	0.947	0.945
	100000	0.952	0.952	0.953	0.952	0.950	0.951	0.948
100	10000	0.860	0.934	0.947	0.946	0.935	0.920	0.851
	100000	0.947	0.950	0.951	0.950	0.947	0.948	0.940
M/M/1 process								
d	n	$\alpha = 0.05$	0.25	0.45	0.65	0.85	0.95	0.99
6	10000	0.949	0.948	0.948	0.948	0.948	0.949	0.947
	100000	0.949	0.952	0.951	0.948	0.947	0.947	0.950
20	10000	0.952	0.950	0.949	0.949	0.949	0.951	0.942
	100000	0.948	0.948	0.948	0.948	0.947	0.949	0.949
100	10000	0.893	0.945	0.950	0.948	0.939	0.918	0.836
	100000	0.943	0.945	0.947	0.945	0.945	0.946	0.942

can observe from the half length expression $d^{-1/2}\hat{\sigma}_{\alpha,d}t_{d-1,1-\gamma/2}$. To this end, as $m \rightarrow \infty$ we have $m(d-1)\hat{\sigma}_{\alpha,d}^2\tau_{\alpha}^2/\sigma_{\alpha,\infty}^2 \Rightarrow \chi_{d-1}^2$. Let $m = n/d$. Then the rescaled half length $\sqrt{nd}^{-1/2}\hat{\sigma}_{\alpha,d}t_{d-1,1-\gamma/2}$ converges in distribution to $t_{d-1,1-\gamma/2}\sqrt{\chi_{d-1}^2/(d-1)\sigma_{\alpha,\infty}/\tau_{\alpha}}$. The latter distribution, which is a rescaled χ distribution, concentrates around and converges to $z_{1-\gamma/2}\sigma_{\alpha,\infty}/\tau_{\alpha}$ as $d \rightarrow \infty$, where $z_{1-\gamma/2}$ is the $(1-\gamma/2)$ th quantile of the standard normal distribution. Nevertheless, a d that is too large can result in poor empirical coverage probabilities as shown in Table 2.9. For example, if sample size is 10,000, $d = 100$ is not a good choice. If d is too big, then the normal approximation for $\hat{\theta}_{\alpha,i}$, which is computed based on n/d observations, can be bad, and the dependence between them can be strong. In general, choosing a smaller batch with a smaller sample is wiser, and vice versa. Similar claims are also made in the literature; see Schmeiser [1982], Law [2013], Munoz et al. [2016] among others. As a rule of thumb, we can choose $d = 2$ if $n \leq 50$, $d = 3$ if $50 < n \leq 100$, and $d = \lfloor n^{1/3} \rfloor$ if $n > 100$. The latter choice of d has an overall satisfactory performance.

CB Simulation

We employ an overlapping block method, as detailed in Politis et al. [2012] and Lahiri [2003], to construct SCBs and PCBs. Unlike the batching approach used for CI estimation in the previous subsection, which is not feasible for computing SCBs, the overlapping block method offers a viable alternative. To facilitate the computation, the algorithm is stated step by step below. Given the data $\{X_1, \dots, X_n\}$ with sample size n , we proceed as follows:

- (i.) Regroup the series $\{X_1, \dots, X_n\}$ into $n-m-1$ consecutive sub-series $\{X_i, X_{i+1}, \dots, X_{i+m-1}\}$, $i = 1, \dots, n-m+1$, each containing m points.
- (ii.) Discretize the given range $[\alpha_1, \alpha_2]$ into N points set $\{\alpha_{N,j} = \alpha_1 + (\alpha_2 - \alpha_1)(j-1)/(N-1) : j = 1, \dots, N\}$.
- (iii.) Based on each block $\{X_i, X_{i+1}, \dots, X_{i+m-1}\}$, calculate the $\alpha_{N,j}$ th expectiles $\hat{\theta}_{\alpha_{N,j},i}$,

$$j = 1, \dots, N.$$

(iv.) Calculate the grand average $\alpha_{N,j}$ th expectiles over all blocks: $\bar{\hat{\theta}}_{\alpha_{N,j}} = \sum_i \hat{\theta}_{\alpha_{N,j},i}/N$,
 $j = 1, \dots, N$.

(v.) Given the confidence level γ , calculate γ th quantile $\hat{q}_\gamma$ of the set $\{\max_j |\hat{\theta}_{\alpha_{N,j},i} - \bar{\hat{\theta}}_{\alpha_{N,j}}| : i = 1, \dots, n - m - 1\}$.

(vi.) The simultaneous confidence band with confidence level γ is constructed as $\hat{\theta}_\alpha \pm \hat{q}_\gamma$,
 $\alpha \in [\alpha_1, \alpha_2]$, where $\hat{\theta}_\alpha$ is the empirical α th expectile of $\{X_1, \dots, X_n\}$.

For comparison purpose, the overlapping block approach is also used for constructing PCBs that is equivalent to calculating CIs for multiple expectile levels.

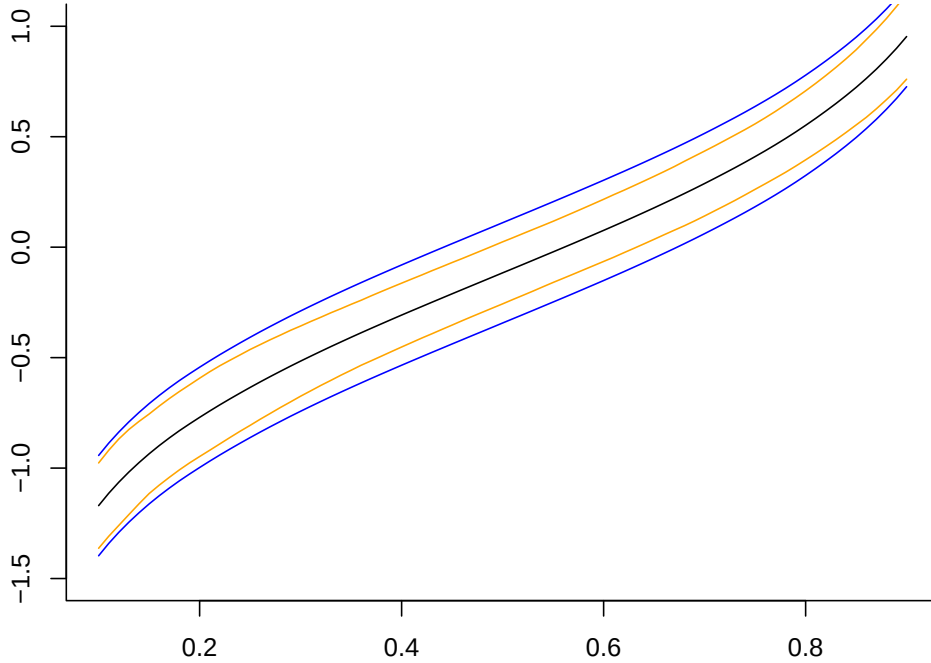


Figure 2.1: SCB in blue vs PCB in yellow with estimated expectiles in black for the AR(1) process with $t(5)$ innovations

Figure 2.1 displays the 95% SCB (blue) and PCB (yellow) for the AR(1) process with $\phi = 0.3$ and $t(5)$ innovations, where α varies from 0.1 to 0.9 in increments of 0.01. The

sample size here is $n = 500$ and the block size is $\lfloor 500^{1/2} \rfloor = 22$. The black line stands for the estimated expectile curve. The figure shows that the SCB covers the PCB, as expected. And based on 10,000 replicates, when $\phi = 0.3$ (resp. $\phi = -0.3$), the empirical coverage probability of the SCB is 92.96% (resp. 94.19%), which are reasonably close to the nominal level of 95%. For other TAR and G/G/1 processes, we obtain similar results.

2.4.2 Expectile Estimation for LRD Processes

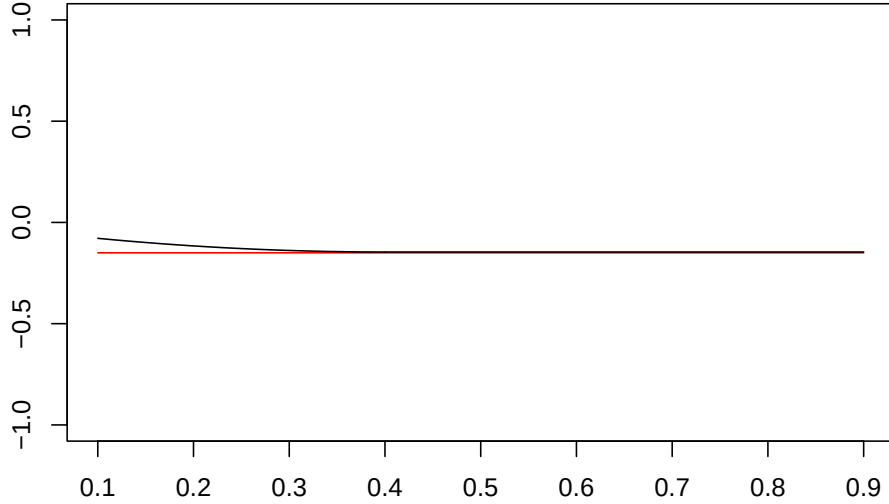


Figure 2.2: Plot of differences $\hat{\theta}_\alpha - \theta_\alpha$, $0.1 \leq \alpha \leq 0.9$ (black) vs $\bar{X}_n$ (red)

This section shows simulations of expectile estimators for long-memory processes. Figure 2.2 visualizes the insights behind (2.22) with one realization of (2.20), where ε_t are i.i.d. $\mathbf{N}(0, 1)$, $a_0 = 1$, $a_n = n^{-\beta}$ for $n \geq 1$ with $\ell(n) = 1$ and $\beta = 0.7$. The black line displays the differences $\hat{\theta}_\alpha - \theta_\alpha$, $0.1 \leq \alpha \leq 0.9$, and the red line represents the sample mean. We can see that the two lines are close, as expected from (2.22). If we apply the d -fold method in Section 3.1.1 to construct pointwise CIs for LRD processes, it completely fails to

achieve the desired coverage probability. We still choose $\mathbf{N}(0, 1)$ innovations and $\beta = 0.7$. For $n = 1000, 5000, 10000$ and $d = 2, 10, 20$, we calculate the probability rates for each combination of (n, d) based on 10,000 replications. Table 2.10 shows that even when $d = 2$, the coverage rates deviate significantly from the nominal level of 95%, regardless of the sample size. Furthermore, as d increases, the coverage rates decline rapidly. These numerical results suggest that the d -fold method designed for SRD processes is not applicable in the LRD case.

Table 2.10: Empirical coverage probability of two-sided 95% CIs of expectiles for the LRD linear process with $\beta = 0.7$ and innovations $\mathbf{N}(0, 1)$ using the d -fold method with sample size n and expectile level α

d	n	$\alpha = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	1000	0.906	0.906	0.911	0.907	0.906	0.908	0.907	0.906	0.905
	5000	0.912	0.911	0.909	0.911	0.911	0.911	0.909	0.909	0.908
	10000	0.910	0.908	0.908	0.908	0.910	0.909	0.908	0.906	0.904
10	1000	0.542	0.545	0.547	0.549	0.551	0.552	0.550	0.550	0.554
	5000	0.573	0.574	0.574	0.573	0.573	0.574	0.576	0.578	0.584
	10000	0.577	0.573	0.572	0.571	0.570	0.571	0.572	0.576	0.578
20	1000	0.419	0.431	0.438	0.441	0.443	0.445	0.444	0.443	0.436
	5000	0.457	0.461	0.465	0.468	0.472	0.476	0.479	0.480	0.485
	10000	0.464	0.467	0.469	0.469	0.470	0.472	0.475	0.478	0.484

The asymptotic result (2.22) suggests the surprising property that SCBs and PCBs are asymptotically identical for long-memory processes. In particular, one can use the CI for any α to construct SCB by re-centering at expectile estimates at other α . The latter property

Table 2.11: Uniform coverage probabilities for AR(1) and LRD processes with $\mathbf{N}(0, 1)$ and $t(3)$ innovations

innovation	n	AR(1)				LRD	
		$\phi = 0$	0.3	0.5	0.9	$\beta = 0.7$	0.55
$\mathbf{N}(0, 1)$	1000	0.797	0.848	0.857	0.871	0.934	0.938
	5000	0.797	0.847	0.863	0.873	0.934	0.935
	10000	0.795	0.847	0.865	0.874	0.936	0.937
$t(3)$	1000	0.467	0.573	0.643	0.783	0.917	0.936
	5000	0.465	0.565	0.619	0.774	0.919	0.936
	10000	0.463	0.568	0.614	0.769	0.921	0.937

is not shared by short-memory processes. For short-memory processes by (2.16), PCBs are not asymptotically close to SCBs. For a comparison study, we examine coverage rates based on the following two settings:

SRD AR(1) process: $X_t = \phi X_{t-1} + \varepsilon_t = \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}$, for $\phi = 0, 0.3, 0.5, 0.9$;

LRD process: $Y_t = \sum_{i=0}^{\infty} a_i \varepsilon_{t-i}$, $a_i = n^{-\beta}$, for $\beta = 0.55, 0.7$.

For each case, we choose three levels of n : 1000, 5000, and 10000 for two types of innovations: $\mathbf{N}(0, 1)$ and $t(3)$. In each run, we generate 10^4 sample paths $\{X_t^{(j)}\}_{t=1}^n$, $j = 1, \dots, 10^4$. For each sample path we compute the sample mean $\bar{X}^{(j)}$. Then we choose cut-off values using sample quantiles based on the sample means $\{\bar{X}^{(j)}\}_{j=1}^{10^4}$ to build SCB of expectiles for each sample path, and calculate the proportion with successful coverage. Table 2.11 shows that the uniform coverage probabilities of long-memory processes are much closer to the nominal level 0.95 than those of AR(1) processes with both $\mathbf{N}(0, 1)$ and $t(3)$ innovations. Additionally, for different β 's in LRD linear process, the coverage rate with smaller β , which indicates stronger dependence, is closer to 0.95 than that with larger β . This rule also works for AR(1) processes: coverage rates with stronger dependence with larger ϕ are closer to the nominal level 0.95.

Figure 2.3 draws SCBs (blue) and PCBs (yellow) for long-memory linear processes with $\beta = 0.55$ and 0.7 , respectively. The SCBs and PCBs are quite close, numerically suggesting that they are asymptotically the same for the LRD process, as (2.22) indicates. Moreover, the width of SCB is about 4% wider than that of PCB for $\beta = 0.55$ while the respective increasing rate is about 6% for $\beta = 0.7$, suggesting that two types of CBs tend to be closer when the dependence is stronger.

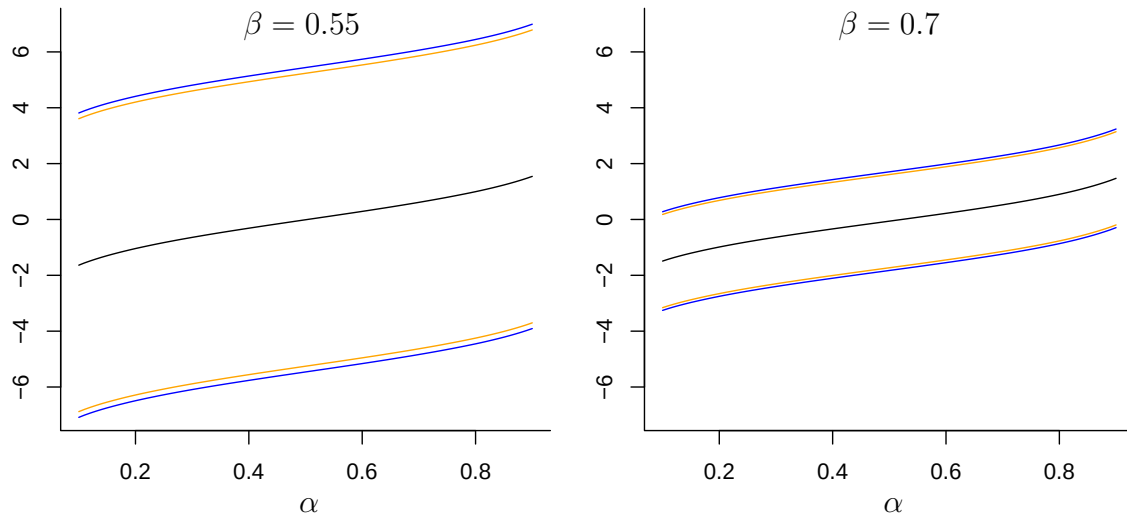


Figure 2.3: SCBs (blue) vs PCBs (yellow) for LRD linear processes with $\beta = 0.55$ and 0.7

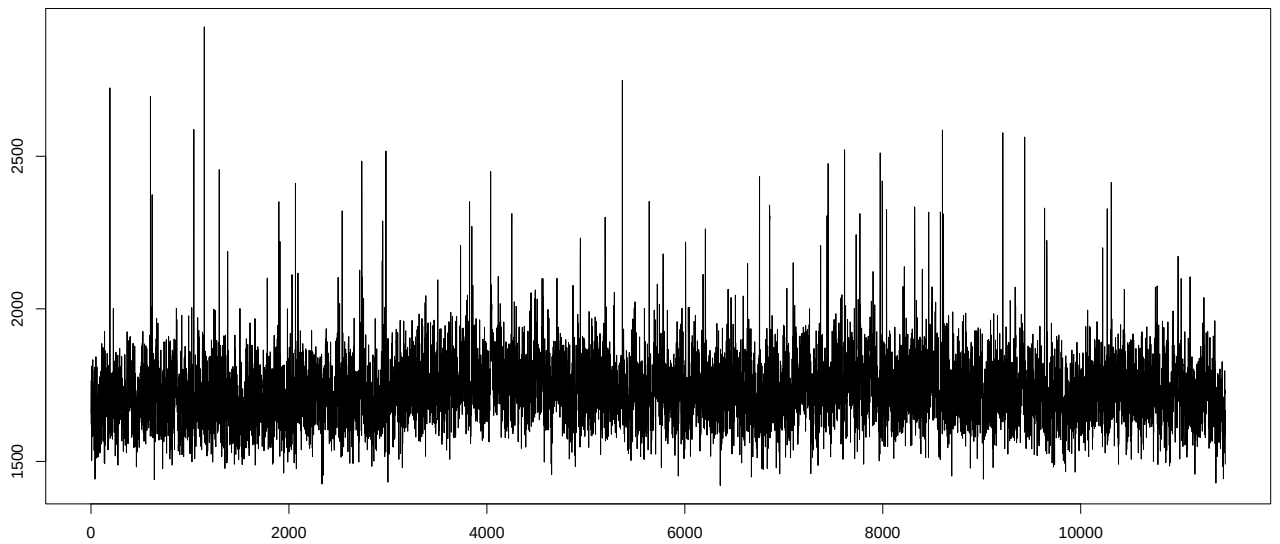


Figure 2.4: Network traffic measured in microseconds, starting at 14:01:00 UTC lasting 11.46 seconds on January 17, 2019

2.5 Application

Network traffic analysis has applications across a wide range of areas and has attracted significant academic interest in monitoring network traffic [Xi et al., 2022], detecting anomalies [Tsimenidis et al., 2022], and predicting volumes [Hu et al., 2024]. Expectiles play a vital role in identifying outliers and abnormal behavior, as they are more sensitive to the tails of the distribution compared to quantiles, making them suitable for network data analysis where extreme events matter.

In this section, we shall analyze a standard and widely studied anonymized internet traces dataset provided by the Center for Applied Internet Data Analysis (CAIDA), a research unit at the University of California San Diego. This data allows us to explore network traffic patterns and uncover overall trends. The file we use is "equinix-nyc.dirB.20190117-140100.UTC.anon.pcap", which contains a total of 19,871,083 packets over the period of 11.46 seconds starting at 14:01:00 UTC on January 17, 2019.

We aggregate 19,871,083 packets with nanosecond unit to microsecond precision, leading to 11,461 observations. Figure 2.4 plots the network traffic time series in microseconds. The autocorrelation function (ACF) test suggests that the SRD condition holds. We can adopt the methods in Section 2.4.1 to compute PCBs and SCBs with different coverage probabilities. CBs are useful in detecting abnormal traffic patterns that deviate significantly from the expected range, such as Distributed Denial of Service (DDoS) attacks or unexpected traffic spikes. Figure 2.5 displays 95% SCB (blue) and PCB (yellow) for network traffic data in the left panel and 99% in the right panel with estimated expectiles (black), which helps to evaluate the potential risk of extreme network delays or packet loss events. By providing a more accurate representation of the tail behavior of traffic distributions, expectile CBs allow network operators optimize resource allocation, considering both typical and worst-case scenarios.

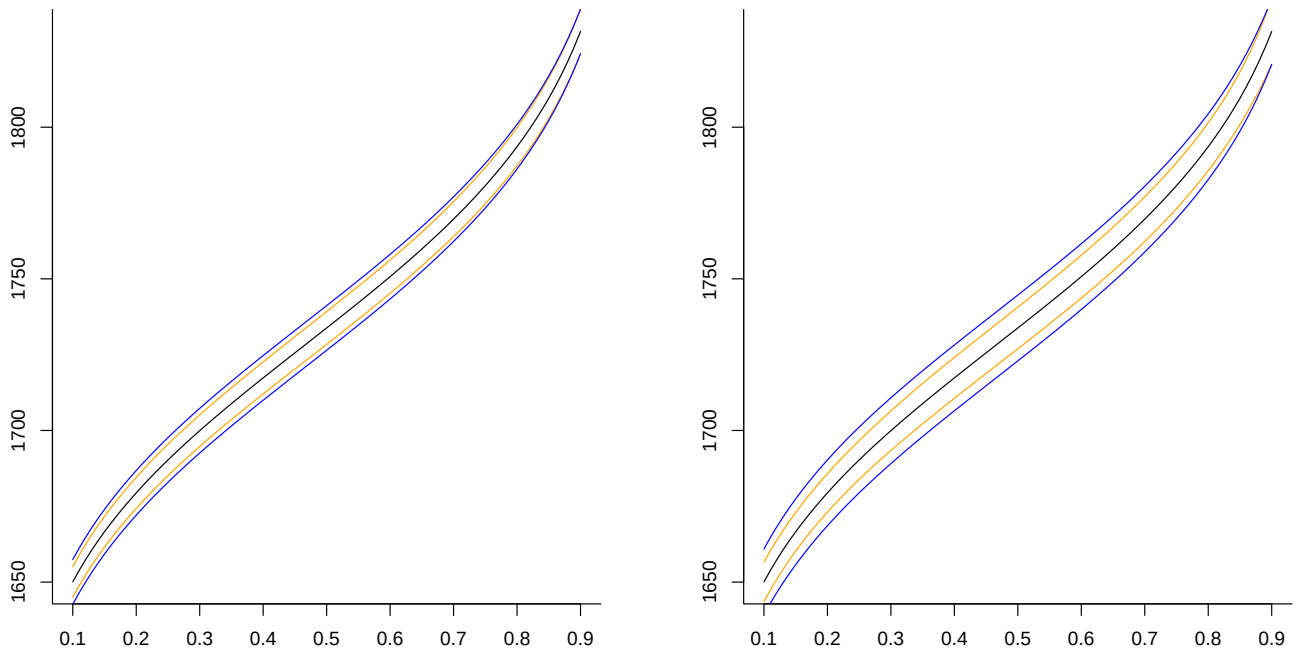


Figure 2.5: SCBs (blue) vs PCBs (yellow) for network traffic data with 95% coverage probability in the left figure and 99% in the right figure

2.6 Conclusion

This paper presents a systematic asymptotic theory of expectile estimators for SRD and LRD processes. For SRD processes, we develop a rigorous asymptotic theory for expectile estimators, making them readily applicable in contexts such as financial time series analysis. In these applications, the dependency structure of the data has often been overlooked, primarily due to a lack of theoretical understanding regarding how such dependencies influence the behavior of expectiles. This gap in theory poses challenges for practitioners seeking to effectively utilize expectiles in the presence of complex dependencies. We first prove a CLT at a single level under the SRD condition using asymptotic linear approximation with Bahadur representation. Our CLT leads to a batching method for constructing CIs for the expectiles. We provide a rule of thumb for selecting batch sizes. It is advisable to select a small batch size when the sample size is small and vice versa. We then develop a weak convergence theory so that SCBs can be constructed for expectiles of SRD processes using a subsampling approach. For LRD processes we also develop an asymptotic theory, which is very different from the one under SRD, and prove the surprising result that SCBs and PCBs are asymptotically identical.

Simulation results support the asymptotic theory. We can see decent performance of empirical coverage probabilities and halfwidths of expectile CIs given the appropriate (n, d) . Based on the overlapping block method, we can construct SCBs for the expectile processes with satisfactory coverage rates and reasonable bandwidths compared to PCBs. For LRD case, we can see that PCBs and SCBs are asymptotically the same. We apply our methods to a network traffic data. It helps to evaluate the potential risk exposure to extreme network service load. We expect that our asymptotic theory for expectile estimators of stationary processes can provide a rigorous foundation for risk management for temporal dependent data.

CHAPTER 3

SIMULTANEOUS INFERENCE OF REGRESSION WITH MULTIVARIATE TIME-VARYING RANDOM COEFFICIENTS

3.1 Introduction

Consider the semi-parametric linear regression model in time series:

$$Y_i = \boldsymbol{\beta}(Z_i)^\top X_i + \sigma(Z_i)\varepsilon_i, \quad i = 1, \dots, n, \quad (3.1)$$

where the coefficient function $\boldsymbol{\beta}(\cdot) = (\beta_1(\cdot), \dots, \beta_p(\cdot))^\top$ in the conditional mean and the conditional standard deviation function $\sigma(\cdot)$ are unknown and need to be estimated. The stationary processes (Y_i, X_i, Z_i) in time series are observed, while the unobserved noise ε_i is an i.i.d. process with $\mathbb{E}\varepsilon_i = 0$ and $\mathbb{E}\varepsilon_i^2 = 1$. The covariate vector X_i is p -dimensional, while Z_i , the argument of the coefficient functions $\boldsymbol{\beta}(\cdot)$ and $\sigma(\cdot)$, is a $(d \times 1)$ random vector. If $p = 2$ and the first coordinate of X_i is one, namely $X_i = (1, x_i)^\top$, then the model (3.1) turns into the widely used simple linear regression with varying coefficients:

$$Y_i = \alpha(Z_i) + \beta(Z_i)x_i + \sigma(Z_i)\varepsilon_i, \quad (3.2)$$

where the intercept $\alpha(\cdot)$ and the slope $\beta(\cdot)$ are scalar functions and the conditional heteroskedasticity is modeled through $\sigma(\cdot)$. The properties of the random variables in (3.2) remain the same as those in (3.1), except for that x_i and $\beta(\cdot)$ are scalars in (3.2). The main purpose of this paper is to illustrate how one can conduct *simultaneous inference* of the coefficient functions in (3.1), namely $\boldsymbol{\beta}(\cdot)$ and $\sigma(\cdot)$.

The regression with varying coefficients in (3.2), more generally in (3.1), is an intuitive approach to take in economics and finance because the relationship among economic (or financial) variables of interest is often highly likely to be dynamic (or state-dependent) rather

than static. By modeling the regression coefficients as nonparametric functions of related random processes as in (3.1) or (3.2), one can investigate the dynamics of the underlying model more effectively, compared to the one with static coefficients. To make this point clear, we consider the celebrated Taylor rule (Taylor [1993]) frequently used for central banks' monetary policies:

$$r_i = \beta_y y_i + \beta_\pi \pi_i + \varepsilon_i, \quad (3.3)$$

where r is the Federal funds rate, π is the inflation rate over the past four quarters and y is the percent deviation of real GDP from a target at time i . It is well-known that the coefficients β_y and β_π are *regime-specific* and that the coefficients vary over time in accordance with underlying state variables, such as the Federal Reserve leadership. That is, (3.3) can be viewed as a specification of our general framework in (3.1). Hence the methodology developed in this study readily applies to (3.3) for more precise and more realistic policy analysis. Other useful applications of (3.1) include the forward premium anomaly or factor pricing models, among others. We refer to Section 3.4 for details on this.

In recent years, much work has been done on estimation and inference of frameworks similar to (3.2) (more generally, to (3.1)): Engle et al. [1986b], Rice [1986], Robinson [1988b], Speckman [1988], Schick [1996], Fan and Huang [2005b] and Xie and Huang [2009] propose the $\sqrt{n}$ -consistent estimators of $\beta(\cdot)$ when $\beta(\cdot)$ is a vector of fixed parameters. The estimation of $\alpha(\cdot)$ under a fixed β has also been intensively studied based on a kernel or spline smoother; see Hamilton and Truong [1997], Yu and Ruppert [2002], Fan and Li [2003] and Aneiros-Pérez and Vilar-Fernández [2008].

Several other studies also perform either point-wise (Liang et al. [1997b]) or simultaneous (Kim et al. [2021b]) inference of $\alpha(\cdot)$ under an i.i.d. setting. Johnston [1982a], Härdle [1989b], Zhao and Wu [2008c], Liu and Wu [2010a], Härdle and Song [2010] estimate and conduct inference of $\alpha(\cdot)$ and/or $\sigma(\cdot)$ under either an i.i.d or a time-dependent setting when $\beta(\cdot) = 0$. Wu and Zhao [2007b] conduct simultaneous inference of $\alpha(\cdot)$ when $\beta(\cdot) = 0$, $\sigma(\cdot) = 1$ and

Z_i is a time index. Kim [2016b] and Li et al. [2025a] extends Wu and Zhao [2007b] to the case where $\beta(\cdot)$ is a vector of fixed parameters. Zhou and Wu [2010a] perform simultaneous inference of the coefficient functions in (3.1) when Z_i is a deterministic time index. Kim and Kim [2016b] apply the methodology in Zhou and Wu [2010a] to a factor pricing model to conduct simultaneous inference of the conditional heteroskedasticity in the model.

Our contribution is to generalize all these aforementioned studies by modeling both $\alpha(\cdot)$ and $\beta(\cdot)$ as unknown functions of the vector process Z_i and also by performing *simultaneous inference* of all the coefficient functions in (3.1). Specifically, the simultaneous inference of (3.1) is conducted through the construction of *simultaneous confidence regions (SCR)* for the coefficient functions. Consider testing a hypotheses for (3.1):

$$H_0 : \beta(\cdot) = \beta_\theta(\cdot) \quad \text{versus} \quad H_1 : H_0 \text{ is not true}, \quad (3.4)$$

where $\beta_\theta(\cdot)$ is a multivariate *parametric* function suggested by related economic theory. To test the hypothesis in (3.4), one can construct the *SCR* of $\beta(\cdot)$ and observe whether the constructed *SCR* *entirely* contains the parametric specification in (3.4). The construction of the *SCR* with confidence level $100(1 - a)\%$, $a \in (0, 1)$, requires us to find two functions $f_n(\cdot)$ and $g_n(\cdot)$ based on data, such that:

$$\lim_{n \rightarrow \infty} \mathbb{P}\{f_n(\mathbf{z}) \leq \beta(\mathbf{z}) \leq g_n(\mathbf{z}) \text{ for all } \mathbf{z} \in \mathcal{Z}\} = 1 - a, \quad (3.5)$$

where $\mathbf{z}$ is a $(d \times 1)$ vector in a compact set $\mathcal{Z}$. Given the *SCR* of $\beta(\cdot)$, one can test (3.4) by checking whether or not the condition $f_n(\mathbf{z}) \leq \beta_\theta(\mathbf{z}) \leq g_n(\mathbf{z})$ holds for *all* $\mathbf{z} \in \mathcal{Z}$. If the condition does not hold for *some* $\mathbf{z} \in \mathcal{Z}$, then we *reject* the null hypothesis at level c . That is, even if the condition holds for all $\mathbf{z} \in \mathcal{Z}$ except for only one, the null hypothesis still gets rejected by the *SCR*-based test.

The relative advantage of our *SCR*-based inference over other standard inferential pro-

cedures utilizing some mean-integrated-squared-error (MISE) type statistic, such as that in Härdle and Mammen [1993], and its associated p -value is in the effectiveness of the SCR -based inference in suggesting the *right* function form of $\beta(\cdot)$ and $\sigma(\cdot)$ in (3.1). In case the null hypothesis in (3.4) is somehow rejected, it would be rather difficult to figure out the reason for rejection if the inference is based on such a MISE-type test statistic and its associated p -value. Thus, it would not be straightforward in such a case to suggest an alternative to the parametric function under the null hypothesis. However, if the inference is based on the proposed SCR , one can easily figure out the reason for the rejection by “locating graphically” where the SCR is violated by the parametric null $\beta_\theta(\cdot)$. Hence it would be relatively straightforward to propose a suitable alternative parametric form for the regression coefficient function.

Roadmap: Section 3.2 introduces our methodology to perform simultaneous inference of $\beta(\cdot)$ in (3.1) with $d \geq 1$. The asymptotic properties of the proposed statistics and estimators are investigated as well. Section 3.3 is devoted to a simulation study to evaluate the performance of our methodology. Section 3.4 introduces an empirical application of testing the currency market efficiency, to demonstrate the applicability/usefulness of the proposed methodology. Section 3.5 concludes the paper and discusses potential extensions for future research. The technical proofs are deferred to Appendix.

Notation: For a vector $\mathbf{v} = (v_1, v_2, \dots, v_p) \in \mathbb{R}^p$, write $|\mathbf{v}| = (\sum_{i=1}^p v_i^2)^{1/2}$. For a random vector $\mathbf{V}$, we write $\mathbf{V} \in \mathcal{L}^q$ ($q > 0$) if $\|\mathbf{V}\|_q = [\mathbb{E}(|\mathbf{V}|^q)]^{1/q} < \infty$. In particular, $\|\mathbf{V}\| = \|\mathbf{V}\|_2$. We write $a_n \asymp b_n$ if $|a_n/b_n|$ is bounded away from 0 and ∞ for all large n .

3.2 Methodology

We work on the inference of any linear combination of the regression functions defined by (3.1). That is, we construct a SCR for $c^\top \beta(\cdot)$ for any deterministic vector c .

3.2.1 Estimation

To estimate the regression coefficients function $\beta(\cdot)$ in (3.1), we conduct the following optimization, among other local-smoothing methodologies:

$$\begin{aligned}\hat{\beta}_b(z) &:= \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^n \left(Y_i - \beta^\top X_i \right)^2 K \left(\frac{z - Z_i}{b} \right) \\ &= \left(\sum_i X_i K \left(\frac{z - Z_i}{b} \right) X_i^\top \right)^{-1} \left(\sum_i X_i K \left(\frac{z - Z_i}{b} \right) Y_i \right),\end{aligned}\quad (3.6)$$

where $K(\cdot)$ is a kernel function and b is a smoothing parameter called the bandwidth. Under suitable conditions, the asymptotic bias of $\hat{\beta}_b(z)$ is of order b^2 . Following Wu and Zhao [2007b], we employ the jackknife-type *bias-corrected* estimate in this study:

$$\tilde{\beta}_b(z) := 2\hat{\beta}_b(z) - \hat{\beta}_{\sqrt{2}b}(z), \quad (3.7)$$

where $\hat{\beta}_{\sqrt{2}b}(z)$ is $\hat{\beta}(z)$ from (3.6) with $\sqrt{2}b$, instead of the original bandwidth b . As Wu and Zhao [2007b] point it out, implementing (3.7) is asymptotically equivalent to using a higher-order kernel function $K^*(\cdot)$:

$$K^*(s) := K(s) - \frac{K(s/\sqrt{2})}{\sqrt{2}}, \quad (3.8)$$

Based on the aforementioned estimates, two sample-based fences for $c^\top \beta$ are required to build a $100(1 - a)\%$ SCR, $a \in (0, 1)$, such that the following condition holds:

$$\lim_{n \rightarrow \infty} \mathbb{P}(\beta_{c,\text{lower}}(z) \leq c^\top \beta(z) \leq \beta_{c,\text{upper}}(z), \forall z \in \mathcal{D}) = 1 - a, \quad (3.9)$$

where z is a $(d \times 1)$ vector in a compact set $\mathcal{D}$.

3.2.2 Construction of SCR

To handle the temporal dependence which is inherent in time series, we first explain the *dependence measure* employed in this study: Assume that $(X_i, Z_i) \in \mathcal{L}^q$, $q > 0$, is a stationary casual process and η_i 's are i.i.d., then one can write

$$\begin{pmatrix} X_i \\ Z_i \end{pmatrix} = G(\xi_i), \text{ where } \xi_i = (\dots, \eta_{i-1}, \eta_i), \quad (3.10)$$

where $G(\cdot) = (g_1(\cdot), \dots, g_p(\cdot), g_{p+1}(\cdot), \dots, g_{p+d}(\cdot))$ is a measurable function such that (X_i, Z_i) exists. Let $\{\eta'_i\}$ be an i.i.d. copy of $\{\eta_i\}$ and

$$\begin{pmatrix} X'_i \\ Z'_i \end{pmatrix} = G(\xi'_i), \text{ where } \xi'_i = (\dots, \eta'_{-1}, \eta'_0, \eta_1, \dots, \eta_i). \quad (3.11)$$

Following Wu [2005b], we define the $\mathcal{L}^\infty$ functional dependence measure:

$$w_{i,q} = \|G(\xi_i) - G(\xi'_i)\|_q. \quad (3.12)$$

Some regularity conditions on the coefficients $\beta(z)$, $\sigma(z)$, (X_i, Z_i) and the bandwidth b_n are required to ensure that the theoretical results in this study hold:

Condition 3.2.1. The regression function $\beta(z) \in \mathcal{C}^3(\mathcal{D})$ satisfies that

$$\sup_{z \in \mathcal{D}} [|\nabla \beta(z)|_\infty + |\nabla^2 \beta(z)|_\infty + |\nabla^3 \beta(z)|_\infty] < \infty,$$

and the noise variance function $\sigma(z) \in \mathcal{C}^3(\mathcal{D})$ are bounded from above and below, i.e. $0 < m \leq \sigma(z) \leq M < \infty$, $\forall z \in \mathcal{D}$ for constants m and M .

Condition 3.2.2. Suppose $(X_1^\top, Z_1^\top)^\top \in \mathcal{L}^q$ with some $q > 2$ and $w_{i,q} = O(\rho^i)$ with $\rho \in (0, 1)$.

Condition 3.2.3. The bandwidth b_n satisfies $b_n \rightarrow 0$ and $nb_n \rightarrow \infty$ as $n \rightarrow \infty$.

Condition 3.2.4. The kernel $K(s) \in \text{Lip}_2$ has support $[-1, 1]^d$, and $\int \tau K(\tau) d\tau$ and $\int \tau^2 K(\tau) d\tau$ are bounded.

Condition 3.2.5. Let $f(z)$ be the marginal density of Z_i . The smallest eigenvalue of $V(z) := \mathbb{E}[X_i X_i^\top | Z_i = z] f(z)$ is uniformly bounded away from 0 for $z \in \mathcal{D}$.

Some comments on Conditions 3.2.1-3.2.5 are noteworthy: The smoothness conditions for the regression function $\beta(z)$ and the bounded variance function $\sigma(z)$ are fairly mild. In Condition 3.2.2, the restriction on (X_i, Z_i) is typically called the Geometric Moment Contraction (GMC). It gives a sufficient dependence condition for the Gaussian approximation in (3.15). Many linear and nonlinear time series models (e.g. ARMA models, ARCH models, GARCH models, bilinear models) satisfy Condition 3.2.2; see Shao and Wu [2007] for the details on this. Condition 3.2.3 on the bandwidth is common in the literature. Condition 3.2.4 on the kernel function is also mild, which is satisfied by many popular kernel functions, such as the Epanechnikov kernel. Under Condition 3.2.5, the process Z_i has a density and the conditional covariance matrix of X_i scaled by density $f(z)$ is uniformly lower bounded over $\mathcal{D}$. This ensures that the estimator in (3.6) be asymptotically well-defined.

Semi-parametric inference of $c^\top \beta(z)$ involves asymptotic analysis of the following quantity:

$$\sum_{i=1}^n w_i(z) \varepsilon_i, \quad (3.13)$$

where $w_i(z)$ are weights depending on X_i and Z_i . The following proposition derived by Bonnerjee et al. [2024] asserts the Gaussian approximation for the linear combination forms:

Proposition 3.2.1. *Let $\varepsilon_1, \dots, \varepsilon_n$ be i.i.d. r.v.s with $\mathbb{E}\varepsilon_i = 0$, $m_q = \mathbb{E}|\varepsilon_i|^q < \infty$, $q > 2$; let $v_i = (v_{i1}, \dots, v_{ip}) \in \mathbb{R}^p$, $1 \leq i \leq n$, be deterministic vectors and*

$$S_n^\varepsilon = \sum_{i=1}^n \varepsilon_i v_i = (S_{n1}, \dots, S_{np})^\top.$$

Then on a richer probability space there are random variables $\phi_i \sim N(0, \text{Cov}(\varepsilon_i))$ such that

$S_n^\phi = \sum_{i=1}^n \phi_i v_i$, which has the same covariance structure as S_n^ε , satisfies

$$\sup_{B \in \mathcal{A}^p} |\mathbb{P}(S_n^\varepsilon \in B) - \mathbb{P}(S_n^\phi \in B)| \leq c (\log(np))^{\frac{5q-6}{4q-2}} \left(m_q \sum_{i=1}^n \max_{j \leq p} \left| \frac{v_{ij}}{\sigma_j} \right|^q \right)^{\frac{1}{2q-1}},$$

where $\mathcal{A}^p$ is the set of hyper-rectangles in $\mathbb{R}^p$ and $\sigma_j^2 = \mathbb{E} S_{n,j}^2$. Here c is a constant that only depends on q .

As discussed above, when we attempt to perform the simultaneous inference based on $\max_z |c^\top \tilde{\beta}(z) - c^\top \beta(z)|$, the weighted sum of the i.i.d. noise ε_i will be the main object to investigate. Under Proposition 3.2.1, one can perform the Gaussian approximation for (3.13) through replacing ε_i by $\phi_i \sim_{i.i.d.} N(0, 1)$. The following theorem states the result in detail:

Theorem 3.2.1. *Assume that Conditions 3.2.1, 3.2.2, 3.2.3, 3.2.4, 3.2.5 hold, $\varepsilon_i \in \mathcal{L}^q$ for some $q > 2$, the weight function $\gamma(z) > m' > 0$ on $\mathcal{D}$, and the bandwidth $b = b_n$ satisfies*

$$(\log(n))^{5q-6} n^{2-q} b^{-qd} \rightarrow 0, \quad n^{\frac{1}{q}-\frac{1}{2}} b^{1-\frac{d}{2}} \log(n) \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

Let $\tilde{w}_{i,b}^c(z) = 2w_{i,b}^c(z) - w_{i,\sqrt{2}b}^c(z)$ where

$$w_{i,b}^c(z) = c^\top \left(\sum_i X_i X_i^\top K \left(\frac{z - Z_i}{b} \right) \right)^{-1} X_i K \left(\frac{z - Z_i}{b} \right) \sigma(Z_i) / \gamma(z). \quad (3.14)$$

Then, on a richer probability space, there exist i.i.d. $\phi_1, \dots, \phi_n \sim N(0, 1)$ such that

$$\begin{aligned} & \sup_{u>0} \left| \mathbb{P}^* \left(\sup_{z \in \mathcal{D}} \left| \frac{c^\top \tilde{\beta}(z) - c^\top \beta(z)}{\gamma(z)} \right| > u \right) - \mathbb{P}^* \left(\sup_{z \in \mathcal{D}} \left| \sum_{i=1}^n \tilde{w}_{i,b}^c(z) \phi_i \right| > u \right) \right| \\ &= O_P \left((\log n)^{\frac{5q-6}{4q-2}} n^{\frac{2-q}{4q-2}} b^{\frac{-qd}{4q-2}} \right), \end{aligned} \quad (3.15)$$

where $\mathbb{P}^*$ is the probability measure conditioning on X_i, Z_i . In particular, the bound in (3.15) converges to 0 if $(\log n)^{5q-6} = o(n^{q-2} b^{qd})$.

Remark 3.2.1. The weight function $\gamma(z)$ plays a role of flexibly scaling the width of the confidence band for each single z . For example, if we choose $\gamma(z) \equiv 1$ for all $z \in \mathcal{D}$, then the simultaneous confidence band that we derive based on the theorem does have the uniform width for all $z \in \mathcal{D}$. If we choose $\gamma(z) = 1/f(z)$, which requires to further assume that $f(z) > m' > 0$ on $\mathcal{D}$ for some constant m' , then the band will have a smaller width at certain z with a larger $f(z)$. Intuitively, for those z 's with a higher density, we tend to obtain more information around them, which would result in more accurate and narrower intervals for these z 's.

In (3.15), the supreme is taken over a continuum $z \in \mathcal{D}$. In practical implementation of Theorem 3.2.1, we can approximate the distribution of

$$\max_{z \in \mathcal{D}} \left| \frac{c^\top \tilde{\boldsymbol{\beta}}(z) - c^\top \boldsymbol{\beta}(z)}{\gamma(z)} \right| \quad (3.16)$$

by $|\sum_{i=1}^n W_i \phi_i|_\infty$ for any choice of $1/n^2$ -net of $\mathcal{D}$. That leads to a feasible approach to pick a suitable cut-off value u^* based on the significance level α so that the following holds:

$$\lim_{n \rightarrow \infty} \mathbb{P}^* \left(\max_{z \in \mathcal{D}} \left| \frac{c^\top \tilde{\boldsymbol{\beta}}(z) - c^\top \boldsymbol{\beta}(z)}{\gamma(z)} \right| > u^* \right) = \alpha. \quad (3.17)$$

Furthermore, one can estimate $\sigma(z)$, the conditional standard deviation function in (3.1), through applying the local-smoothing methodology to the least squares residuals:

$$\hat{\sigma}^2(z) := \left(\sum_{k=1}^n K \left(\frac{Z_k - z}{h_n} \right) \right)^{-1} \left(\sum_{i=1}^n K \left(\frac{Z_i - z}{h_n} \right) [Y_i - \tilde{\boldsymbol{\beta}}(Z_i)^\top X_i]^2 \right), \quad (3.18)$$

where the bandwidth $h_n \asymp b_n$. The following theorem provides the asymptotic basis for the construction of a *SCR* for the conditional standard deviation function $\sigma(\cdot)$:

Theorem 3.2.2. *Besides the conditions in Theorem 3.2.1, we further assume*

$$\frac{\log n}{n^{1-2/q}h_n^d} + \frac{1}{n^2h_n^{5d}} \rightarrow 0. \quad (3.19)$$

Then,

$$\sup_{z \in \mathcal{D}} |\hat{\sigma}^2(z) - \sigma^2(z)| = O_P(\tau_n), \quad (3.20)$$

where

$$\tau_n = h_n + \frac{1}{nh_n^{5d/2}} + \frac{\log n}{n^{1-2/q}h_n^d}.$$

The simultaneous inference of the conditional standard deviation function $\sigma(\cdot)$ in (3.1) can be conducted as in Liu and Wu [2010a]. After replacing $\gamma(z)$ in (3.17) by $\hat{\gamma}(z)$, one can construct a $100(1 - a)\%$ SCR, $a \in (0, 1)$, for $c^\top \beta(z)$ such that:

$$c^\top \tilde{\beta}(z) - u^* \hat{\gamma}(z) \leq c^\top \beta(z) \leq c^\top \tilde{\beta}(z) + u^* \hat{\gamma}(z), \text{ for all } z \in \mathcal{D}, \quad (3.21)$$

where the statement holds with an asymptotic probability of $1 - a$.

3.2.3 Multiplier bootstrap

In the previous section we presented a feasible approach to build the SCR for $c^\top \beta(z)$ as shown in (3.21). In practice, however, the discretization of the compact domain $\mathcal{D}$ in (3.21) could cause a high degree of computational complexity. This potential issue could make the proposed methodology intractable. To this end, we consider the *multiplier bootstrap* methodology to handle the potential computational issue. With the estimate $\hat{\sigma}(z)$ in (3.18), one can generate a new response set

$$Y_i^* = \hat{\sigma}(Z_i) \phi_i, \text{ where } \phi_i \sim N(0, 1) \text{ i.i.d.} \quad (3.22)$$

Replacing Y_i by $Y_i^\star$ in (3.6) and (3.7), we can get $\tilde{\beta}_b^\star(z) = 2\hat{\beta}_b^\star(z) - \hat{\beta}_{\sqrt{2}b}^\star(z)$, which is an asymptotically unbiased estimate of zero. The following theorem states that one can approximate the distribution of (3.16) and choose a proper cut-off value $u^\star$ based on $Y_i^\star$.

Theorem 3.2.3. *Assume that the conditions for Theorem 3.2.1 and for Theorem 3.2.2 hold. Then,*

$$\begin{aligned} & \sup_{u>0} \left| \mathbb{P}^* \left(\max_{z \in \mathcal{D}} \left| \frac{c^\top \tilde{\beta}(z) - c^\top \beta(z)}{\gamma(z)} \right| > u \right) - \mathbb{P}^* \left(\max_{z \in \mathcal{D}} \left| \frac{c^\top \tilde{\beta}^\star(z)}{\gamma(z)} \right| > u \right) \right| \\ &= O_P \left((\log n)^{\frac{5q-6}{4q-2}} n^{\frac{2-q}{4q-2}} b^{\frac{-qd}{4q-2}} + \tau_n \right). \end{aligned} \quad (3.23)$$

Theorem 3.2.3 has an important application that one can obtain cutoff values via extensive simulations. Specifically, one can repeat (3.22) for a numerous times (say, $s = 10^4$) to obtain a set of values

$$\left\{ \max_{z \in \mathcal{D}} \left| \frac{c^\top \tilde{\beta}_1^\star(z)}{\hat{\gamma}(z)} \right|, \dots, \max_{z \in \mathcal{D}} \left| \frac{c^\top \tilde{\beta}_s^\star(z)}{\hat{\gamma}(z)} \right| \right\}. \quad (3.24)$$

We then can estimate the cut-off value $u^\star$ in accordance with the quantiles of (3.24). That is, the sampling distribution of $\max_{z \in \mathcal{D}} |c^\top \tilde{\beta}^\star(z)/\gamma(z)|$ based on the simulated $Y_i^\star$ in (3.22) can be utilized to obtain $u^\star$ required for the *SCR* in (3.21).

3.3 Simulation

To illustrate the finite-sample performance of the proposed simultaneous confidence region (*SCR*) based inference, we shall carry out a simulation study here. To that end, we consider the following data-generating process (DGP), a specification of (3.1), to compute the empirical *coverage probability* of the proposed *SCR* for the regression coefficient functions:

$$Y_i = \alpha(Z_{1i}, Z_{2i}) + \beta(Z_{1i}, Z_{2i})X_i + \sigma(Z_{1i}, Z_{2i})\varepsilon_i, \quad i = 1, \dots, n, \quad (3.25)$$

where ε_i are independent and identically distributed (i.i.d.) Student- t_8 . To model its temporal dependence and potential autocorrelation among the variables, we assume a vector autoregressive (VAR) process of order one for the bi-variate Z_{1i} and Z_{2i} :

$$\begin{aligned} \text{VAR}(1) \text{ process} \quad : \quad & Z_{1i} = 0.1 Z_{1i-1} - 0.2 Z_{2i-1} + v_{1i}, \\ & Z_{2i} = 0.05 Z_{1i-1} - 0.1 Z_{2i-1} + v_{2i}, \end{aligned}$$

where v_{1i} and v_{2i} are i.i.d. standard normal processes, independent of ε_i . It is well-known that a VAR specification is a flexible way to model a multivariate stationary process in time series. In addition, the specification is covered by the recursive dependence structure introduced in Condition 3.2.2.

For the covariate term X_i in (3.25), we consider the following three distinct DGPs:

$$\begin{aligned} \text{DGP 1} \quad : \quad & X_i = u_i \\ \text{DGP 2} \quad : \quad & X_i = 0.2 + 0.1 Z_{1i} - 0.3 Z_{2i} + u_i \\ \text{DGP 3} \quad : \quad & X_i = 0.1 - 0.2 Z_{1i}^2 + 0.05 Z_{2i}^2 + u_i \end{aligned}$$

where u_i are i.i.d. standard normals, independent of ε_i and v_i . Note that DGPs 1 and 2 are linear in (Z_{1i}, Z_{2i}) while DGP 3 is non-linear in (Z_{1i}, Z_{2i}) , which helps generalize the simulation outcomes here. Regarding the coefficient functions $\alpha(\cdot, \cdot)$, $\beta(\cdot, \cdot)$ and $\sigma(\cdot, \cdot)$ in (3.25), we employ the following continuous null functions¹:

$$\begin{aligned} \alpha(z_1, z_2) &= 0.1 - 0.4 z_1^2 - 0.05 z_2, \\ \beta(z_1, z_2) &= 0.1 + 0.1 z_1 - 0.05 z_2, \\ \sigma(z_1, z_2) &= \left(0.1 + 0.1 z_1^2 + 0.1 z_2^2 \right)^{1/2}. \end{aligned}$$

1. The empirical results in Section 3.4 indicate that the chosen null functions are reasonable.

For the simultaneous inference of the coefficient functions, we follow the procedures described to construct the 95% *SCR* of $\alpha(\cdot, \cdot)$ and $\beta(\cdot, \cdot)$ in (3.25) under the VAR process of (Z_{1i}, Z_{2i}) and the three DGPs of X_i . The entire process is repeated, say M times, for each one of the chosen bandwidths in Table 3.1. Given M different *SCRs*, we count how many of those *SCRs* *entirely* contain the true $\alpha(\cdot, \cdot)$ and $\beta(\cdot, \cdot)$ in them, which gives us the desired coverage probabilities. Here we let $M = 1,000$ and $n = 300$ (i.e. the sample size).² The simulation results for DGP 1–DGP 3 under different bandwidths are summarized in Table 3.1. Throughout the simulation in this study, we let the nominal coverage rate be 95 percent.

From Table 3.1, we see that the coverage probabilities for $\alpha(\cdot, \cdot)$ and $\beta(\cdot, \cdot)$ under the various bandwidths are reasonably close to the nominal coverage rate 0.95 for each combination of the DGPs for X_i and (Z_{1i}, Z_{2i}) . Despite the relatively small size of sample (i.e. $n=300$) throughout this simulation, the coverage probabilities for both $\alpha(\cdot, \cdot)$ and $\beta(\cdot, \cdot)$ under the various DGPs are fairly close to 0.95. These findings indicate that the proposed methodology does work well. However, as the bandwidth gets extreme for each DGP, the empirical coverage probability deviates from the nominal coverage rate 0.95 for both the intercept term and the slope coefficient, as we can observe from Table 3.1.

The last row of Table 3.1 also reports the 95% coverage probabilities for the coefficient functions under the bandwidth chosen by the generalized cross-validation (GCV). That is, instead of fixing the bandwidth arbitrarily, the GCV is applied to choose the optimal bandwidth for *each* iteration. The construction of a simultaneous confidence region is conducted in each iteration under the GCV-chosen bandwidth to decide whether or not the parametric null function is entirely contained by the simultaneous confidence region. The bottom row in Table 3.1 shows that the coverage probabilities under the GCV-chosen bandwidth are very close to the nominal 95% coverage rate.

2. Due to the slow data-updating process and the occasional structural breaks, the typical sample size in the empirical macro-finance literature is not large. For example, the sample size in Section 3.4, the application section of the current paper, is only $n = 336$.

Table 3.1: 95% coverage probability of SCR for $\alpha(\cdot, \cdot)$ and $\beta(\cdot, \cdot)$, where Z_{1i} and Z_{2i} follow a VAR(1)

bandwidth	$X_i = u_i$		$X_i = 0.2 + 0.1 Z_{1i} - 0.3 Z_{2i} + u_i$		$X_i = 0.1 - 0.2 Z_{1i}^2 + 0.05 Z_{2i}^2 + u_i$	
	$\alpha(\cdot, \cdot)$	$\beta(\cdot, \cdot)$	$\alpha(\cdot, \cdot)$	$\beta(\cdot, \cdot)$	$\alpha(\cdot, \cdot)$	$\beta(\cdot, \cdot)$
0.20	0.982	0.983	0.983	0.982	0.980	0.981
0.22	0.981	0.981	0.980	0.984	0.976	0.981
0.24	0.980	0.976	0.975	0.976	0.970	0.975
0.26	0.973	0.969	0.966	0.964	0.961	0.968
0.28	0.968	0.964	0.964	0.962	0.961	0.965
0.30	0.955	0.957	0.949	0.958	0.952	0.956
0.32	0.953	0.950	0.943	0.952	0.948	0.954
0.34	0.946	0.937	0.935	0.939	0.948	0.931
0.36	0.938	0.937	0.936	0.939	0.936	0.934
0.38	0.925	0.921	0.931	0.930	0.934	0.927
0.40	0.925	0.923	0.926	0.928	0.925	0.929
GCV	0.944	0.948	0.949	0.954	0.943	0.952

3.4 Application: Currency Market Efficiency

The methodology in this work is applied to test the rational expectations hypothesis in the currency market. Fama [1984] and Bilson [1981] test the hypothesis based on the following regression:

$$s_{t+1} - s_t = \alpha + \beta(f_{1,t} - s_t) + u_{t+1} \quad (3.26)$$

where s_t is log of monthly *spot* exchange rate at time t and $f_{1,t}$ is log of monthly *forward* exchange rate with one-month maturity at time t , respectively. The model error u_t could be serially correlated if the maturity of the involved forward contract is longer than the sampling interval; see Baillie et al. [2023]. By Covered Interest Parity, (3.26) yields the Uncovered Interest Parity (*UIP*) regression, where the spot return is predicted by the interest rate differential between two countries.

A large body of the literature on the regression in (3.26) shows that the constant term α , the foreign exchange risk premium, and the UIP coefficient β are closely related to *fundamentals* (Hodrick [1989]; Mark [1995]; Alvarez et al. [2009]; Lustig et al. [2014]; Baillie and Kim [2015a]; Kim et al. [2023]; Li et al. [2025a]), such as the differential volatility in the bond market, conditional variance of money growth, output growth, etc. In view of these works, one can possibly model α and β in (3.26) as unknown functions of the fundamentals:

$$s_{t+1} - s_t = \alpha(z_{1,t}, z_{2,t}) + \beta(z_{1,t}, z_{2,t})(f_{1,t} - s_t) + \sigma(z_{1,t}, z_{2,t})\varepsilon_{t+1}, \quad (3.27)$$

where

$$z_{1,t} := m_t - m_t^*, \quad z_{2,t} = y_t - y_t^*.$$

Here m_t and y_t are log of domestic money stock and log of domestic industrial production, respectively, while m_t^* and y_t^* are their foreign counterparts.³ Note here that α and β in

3. The Augmented Dickey-Fuller test and the Perron-Phillips test based on our data reject the unit-root non-stationary hypothesis for the dependent and independent variables in (3.27).

(3.26) are replaced by $\alpha(z_{1,t}, z_{2,t})$ and $\beta(z_{1,t}, z_{2,t})$, respectively, and that the conditional heteroscedasticity is introduced in (3.27).

The theory of Uncovered Interest Parity (*UIP*) in international finance implies $\alpha(\cdot) = 0$ and $\beta(\cdot) = 1$, while a number of previous empirical studies actually show that the equalities do not hold in practice.⁴ Obviously, the methodology developed in this work can be readily applied to test whether or not the *UIP* conditions hold here, because (3.27) is nothing but a specification of the general framework in (3.1).

Our data include monthly spot rate and 30-day forward exchange rate data for a pair of currencies: Swiss Franc (*CHF*)–U.S. Dollar (*USD*), where *USD* is the numeraire currency. For the money and the output variables, we employ monthly M1, the standard measure of a money supply and monthly industrial production index home (i.e. *U.S.*) and abroad (i.e. Switzerland).⁵ The sample period is from December 1988 through December 2016, including the 2008 financial crisis.

Fig.3.1 illustrates the nonparametric estimates of the coefficient functions $\alpha(\cdot, \cdot)$ and $\beta(\cdot, \cdot)$ in (3.27), the nonlinear surfaces, for the *CHF-USD* pair. The bandwidths are chosen by the generalized cross-validation (*GCV*). As we can see from Fig.3.1, the estimates of both coefficient functions are *highly nonlinear* in the relative money growth. The seeming nonlinearity indicates that (3.26) with no dynamics in α and β is possibly an unsuitable framework for the true processes.

Fig.3.2 shows the two-dimensional relationship between the intercept $\alpha(\cdot, \cdot)$ and the relative money for *CHF/USD* when the relative output is fixed at some percentile. In contrast, Fig.3.3 shows the relationship between $\alpha(\cdot, \cdot)$ and the relative output for *CHF/USD* when the relative money is fixed at a certain percentile.

Similarly, Fig.3.4 shows the relationship between the *UIP* coefficient $\beta(\cdot, \cdot)$ and the rel-

4. Froot and Thaler [1990a] provide an excellent review of the literature on this.

5. Note that the usual GDP series cannot be used here because the sampling frequency is monthly in this application and the highest sampling frequency for GDP data is quarterly.

ative money for CHF/USD at a certain relative output percentile, while Fig.3.5 illustrates the relationship between $\beta(\cdot, \cdot)$ and the relative output at a certain relative money percentile. In addition, each panel in Figs 3.2–3.5 includes the 95% *SCR* of the corresponding coefficient function. The horizontal line represents the *UIP* null hypothesis of $\alpha(\cdot, \cdot) = 0$ in Figs 3.2 and 3.3 and of $\beta(\cdot, \cdot) = 1$ in Figs 3.4 and 3.5.

Interestingly, Fig.3.4 shows the increasing pattern in the *UIP* coefficient $\beta(\cdot)$ estimates. Fama [1984] shows that β can be decomposed into:

$$\beta = \frac{Cov(P_t, \mathbb{E}_t(\Delta s_{t+1})) + Var(\mathbb{E}_t(\Delta s_{t+1}))}{Var(f_t - s_t)},$$

where $\Delta s_{t+1} = s_{t+1} - s_t$ and P is the risk premium. Intuitively, with an relative increase in the money home, the expected spot return and the risk premium of holding a foreign currency would decrease simultaneously, because the foreign currency would appreciate.⁶ The simultaneous decrease in the expected spot return and the risk premium would increase $Cov(P_t, \mathbb{E}_t(\Delta s_{t+1}))$. Assuming that the variance of the expected spot return and that of the forward premium are relatively stable, the increase in the covariance would increase β . That is, a relative increase in the domestic money would lead to an increase in β as shown in Fig.3.4, which is confirmed by Fig.3.5 as well.

Furthermore, Fig.3.4 also shows an interesting change in the β coefficient estimate across the displayed percentiles of the relative output. In Fig.3.4, the estimate of the β coefficient function shifts slightly *downward* as the percentile of the relative output growth increases. It is frequently observed in various empirical studies that currencies with relatively higher output growth tend to possess relatively lower β 's than those with relatively lower output growth; see Kim et al. [2023] for example. The seeming negative relationship between the output growth and the β appears to be captured by Fig.3.4, so that the β estimate (i.e.

6. The intuitive negative relationship between the money and the risk premium (i.e. $-\alpha$ by definition) is confirmed by the positive relationship between α and the money in Figs 3.2 and 3.3.

the solid curve in the figure) shifts slightly down as the percentile of the output growth increases. This negative relationship between the output growth and β is obviously confirmed by Fig.3.5.

In order to accept the null hypothesis of UIP at a 5% level, the horizontal lines must be *entirely* contained by *all* *SCRs* in Figs 3.2–3.5. Obviously, Panels (a)–(c) in Figs 3.2 and Panel (a) of Fig.3.3 fail to contain the horizontal lines entirely. Interestingly, even the general null of a *constant* risk premium, not that of a zero risk premium, is still rejected at a 5% level because no horizontal line can be entirely covered by the 95% *SCRs* in Fig.3.2. Similarly, the null of the unit UIP coefficient is rejected at 5% since the *SCRs* in the panels of Figs 3.4 and 3.5 fail to entirely contain the unit β . In view of Figs 3.2–3.5, the null hypothesis of UIP is thus *rejected at a 5% level*, which is one of the key implications in our study.

The key reason for the UIP rejection is because our estimates of $\alpha(\cdot, \cdot)$ and of $\beta(\cdot, \cdot)$ vary “non-linearly” in the relative money growth. That is, $\alpha(\cdot, \cdot)$ and $\beta(\cdot, \cdot)$ are *not invariant* in their arguments. This finding highlights the relative advantage of our *SCR*-based inference because, with the usual p -value of a *MISE*-type test statistic, it is hard to figure out why the associated null hypothesis gets rejected. Under our approach, however, one can easily locate *exactly where* the *SCR* fails to contain the null hypothesis. Hence it would be relatively straightforward under our methodology to propose an alternative function that fits entirely into the corresponding *SCR*.

3.5 Conclusion

In this study, we introduce a methodology to carry out simultaneous inference of a linear regression in time series when the regression coefficients and the volatility are unknown functions of a *multivariate random* process. To this end, the high-dimensional Gaussian approximation theory (Chernozhuokov et al. [2022]) is generalized to construct the simultaneous confidence region for the multivariate unknown coefficient functions in a time series

regression. In addition, this work is a generalization of the two-dimensional simultaneous confidence band in time series (Zhou and Wu [2010a]; Kim and Kim [2016b]) in that the argument of the coefficient functions is extended to be a *multivariate random* process in our framework. The generalization is not only technically challenging, but also practically useful.

The asymptotic theories justifying the methodology are provided and the related finite-sample properties are examined through a simulation study. The methodology developed is applied to test the hypothesis of rational expectations for the currency market. For this objective, a Swiss Franc–US Dollar currency pair is employed to find out that the rational expectations hypothesis for the currency market is rejected at a 5% level, largely because of the nonlinear nature of the underlying dynamics.

The current study can be extended in a couple of interesting directions: First, the covariate terms and the argument of the coefficient functions could be modeled as some form of *non-stationary* processes. This would significantly generalize the findings in our current work. Second, the innovation could be modeled as a *long memory* process. Persistence is one of the salient features of financial variables, such as the return volatility. By allowing the model innovation to be a process with long memory, we can make our methodology applicable to a wide range of interesting issues in economic/financial analysis. Further insight can be gained by extending the current work in these and other directions.

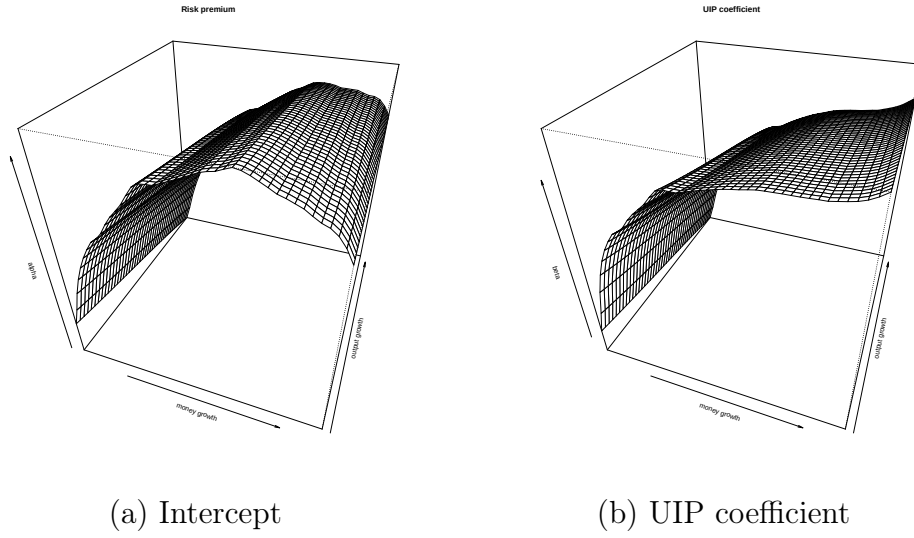


Figure 3.1: Intercept (*i.e.* $\alpha(\cdot)$) and UIP coefficient (*i.e.* $\beta(\cdot)$) in the relative money growth and the relative output growth for **Swiss Franc (CHF)**–**US Dollar (USD)**; The band-width is obtained through the generalized cross-validation (*GCV*).

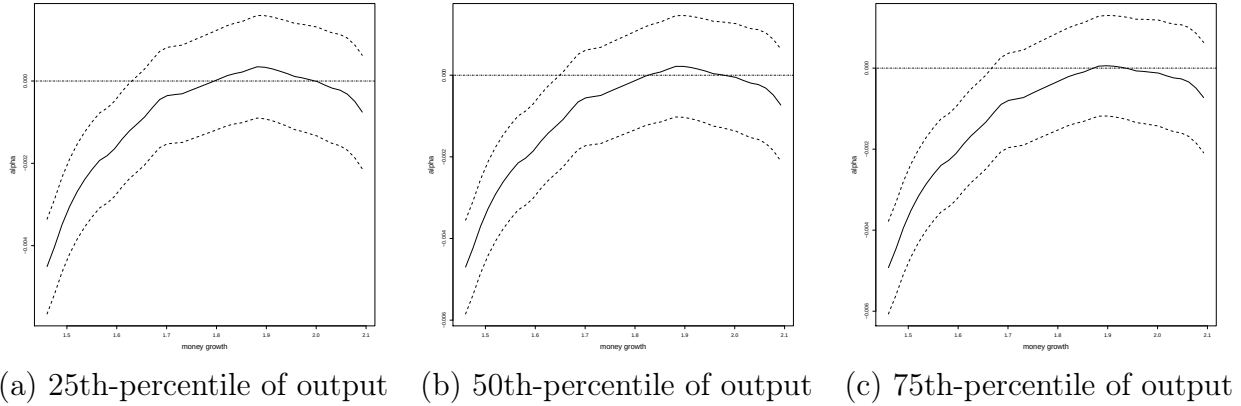


Figure 3.2: The *solid* curve is the local-linear regression estimate of $\alpha(\cdot)$ in relative money growth for **CHF/USD**. The *dashed* band is the 95% *SCR* of $\alpha(\cdot)$ and the dotted horizontal line is $H_0 : \alpha(\cdot) = 0$. The bandwidth is chosen by *GCV*. Here (a)–(c) are 25th, 50th and 75th percentile of output growth, respectively.

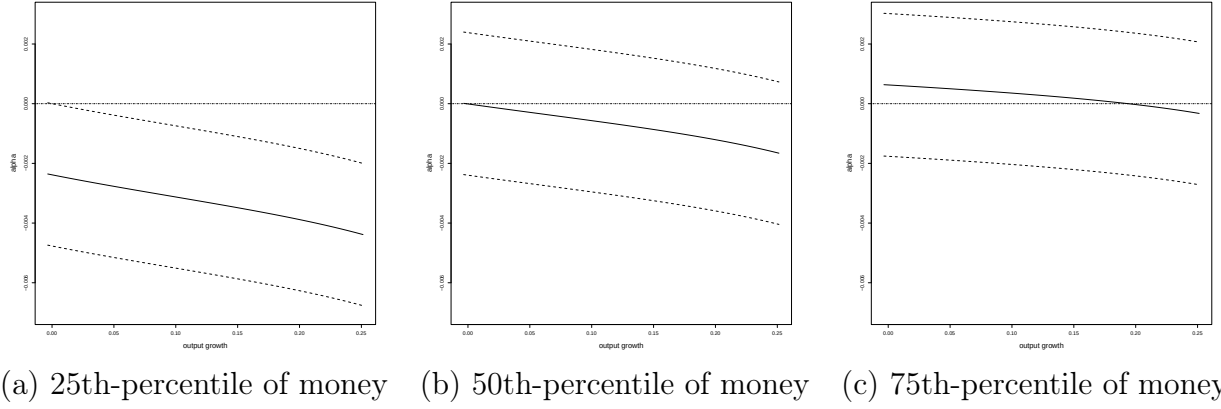


Figure 3.3: The *solid* curve is the local-linear regression estimate of $\alpha(\cdot)$ in relative output growth for **CHF/USD**. The *dashed* band is the 95% *SCR* of $\alpha(\cdot)$ and the dotted horizontal line is $H_0 : \alpha(\cdot) = 0$. The bandwidth is chosen by *GCV*. Here (a)–(c) are 25th, 50th and 75th percentile of money growth, respectively.

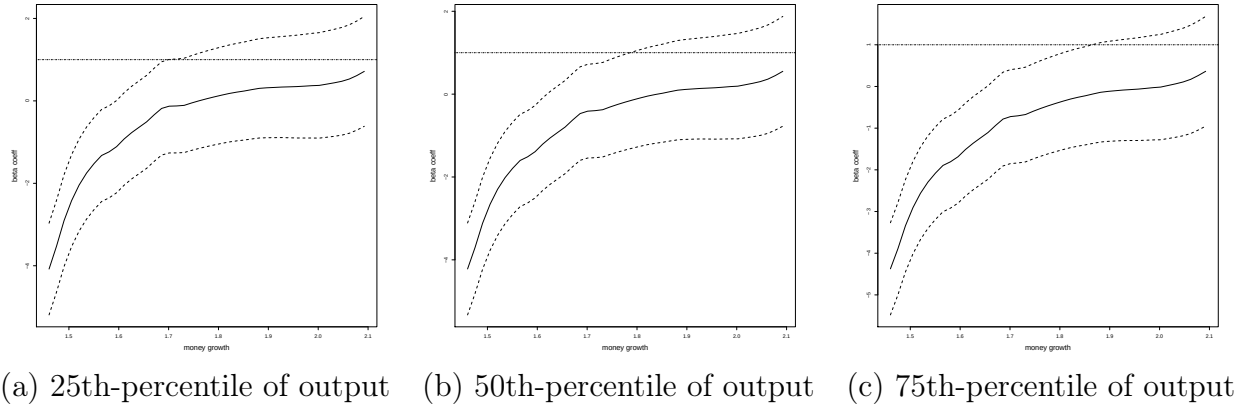
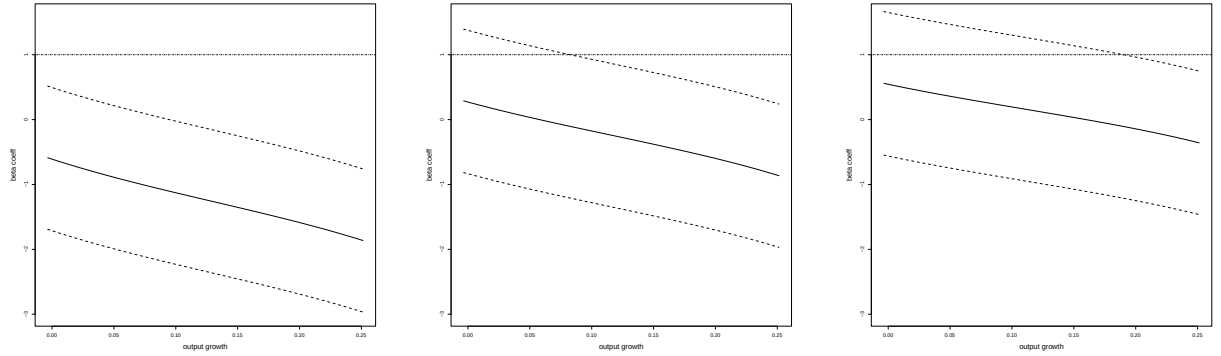


Figure 3.4: The *solid* curve is the local-linear regression estimate of $\beta(\cdot)$ in relative money growth for **CHF/USD**. The *dashed* band is the 95% *SCR* of $\beta(\cdot)$ and the dotted horizontal line is $H_0 : \beta(\cdot) = 1$. The bandwidth is chosen by *GCV*. Here (a)–(c) are 25th, 50th and 75th percentile of output growth, respectively.



(a) 25th-percentile of money (b) 50th-percentile of money (c) 75th-percentile of money

Figure 3.5: The *solid* curve is the local-linear regression estimate of $\beta(\cdot)$ in relative output growth for **CHF/USD**. The *dashed* band is the 95% *SCR* of $\beta(\cdot)$ and the dotted horizontal line is $H_0 : \beta(\cdot) = 1$. The bandwidth is chosen by *GCV*. Here (a)–(c) are 25th, 50th and 75th percentile of money growth, respectively.

CHAPTER 4

HIGH CONFIDENCE LEVEL INFERENCE FOR QUANTILE REGRESSION

4.1 Introduction

Consider the following regression model:

$$Y_i = X_i^\top \boldsymbol{\beta} + e_i, \quad i = 1, \dots, n,$$

where X_i and Y_i are observed covariates and response respectively, $\boldsymbol{\beta} \in \mathbb{R}^p$ is the fixed p -dimensional coefficient vector and e_i are independent and identically distributed (i.i.d.) errors. Consider the quantile regression model at level $\alpha \in (0, 1)$ and we denote X and Y as population version of covariates and response. Then the conditional quantile function of Y is

$$Q_Y(\alpha|X) = X^\top \boldsymbol{\beta} + F_e^{-1}(\alpha),$$

where $F_e(\cdot)$ is the common distribution function of errors. We can estimate $\boldsymbol{\beta}$ with the following M-estimator

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^n \rho_\alpha(Y_i - X_i^\top \boldsymbol{\beta}),$$

where $\rho_\alpha(x) = \alpha x^+ + (1 - \alpha)(-x)^+$ and $x^+ = \max(x, 0)$. Let $\boldsymbol{\beta}_0$ be the true value of $\boldsymbol{\beta}$. Suppose we are able to construct asymptotic $(1 - \tau)$ -confidence regions $\hat{\mathcal{R}}_n$ for $\boldsymbol{\beta}$, which means

$$\mathbb{P}(\boldsymbol{\beta} \in \hat{\mathcal{R}}_n) \approx 1 - \tau.$$

The commonly used confidence level would be 95% or 90% in various disciplines, such as zoology, climate science and medicine production. In this case, many literature have developed Bahadur representation and central limit theorem for $\hat{\boldsymbol{\beta}}$ to conduct inference with

large sample size (see Koenker [2005], Wu [2007b]). However, when τ is typically small, like $\tau = 10^{-3}$, the inference becomes quite nontrivial, where high confidence level issue emerges. Empirically, a tremendous sample size would be desired to construct a .999-confidence region. Another potential issue would be that, as the confidence level goes extremely close to 1, the difference of region performance becomes hard to distinguish with respect to the absolute error. For example, a .999-confidence region $\mathcal{R}_1$ and .99-confidence region $\mathcal{R}_2$ only differ 0.009 in the confidence level, but the previous one is 10 times better than the latter one, because

$$\frac{\mathbb{P}(\boldsymbol{\beta} \notin \mathcal{R}_2)}{\mathbb{P}(\boldsymbol{\beta} \notin \mathcal{R}_1)} = 10.$$

To this end, we desire to develop a moderate deviation principle for $\hat{\boldsymbol{\beta}}$ so that the following bound could be derived:

$$|\mathbb{P}(\sqrt{n}|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0| > ca_n) - \mathbb{P}(|Z| > ca_n)| \rightarrow 0, \quad (4.1)$$

where Z is normal, $c > 0$ is a constant, and $\{a_n\}$ is a positive sequence such that $a_n \rightarrow \infty$ and $n^{-1/2}a_n \rightarrow 0$ as $n \rightarrow \infty$. With such kind of result, we can ensure the convergence on the extreme tails. Moreover, if the bound in 4.1 is sharp enough, we can derive that the relative error also converges to 0. While Moderate deviation principles have been extensively studied in mathematical statistics (see Amosova [1974], Rubin and Sethuraman [1965], Liming [1995], Melbourne [2009], Wu [2001]), the moderate deviation of quantile regression coefficients is still an open question. We will show the theoretical results in section 4.2.

One application is testing the trend of coefficients for longitudinal data in the following model:

$$Y_{t,i} = X_{t,i}^\top \boldsymbol{\beta}_t + e_{t,i}, \quad i = 1, \dots, n_t, \quad t = 1, \dots, T.$$

There are some influential work on the trending test. Wu and Zhao [2007c] conduct simultaneous inference of a trend when the argument of the trend function is simply a time index.

Kim [2016a] extends Wu and Zhao [2007c] to the case where the trend in a semi-parametric partially linear model is a smooth function of a time index. Zhou and Wu [2010b] perform simultaneous inference of the coefficient functions in a regression where the argument of a non-parametric regression coefficient function is a deterministic time index. Baillie and Kim [2015b] applies Zhou and Wu [2010b] to the forward premium regression to show that the foreign exchange risk premia not only depend on macroeconomic fundamentals but also can be predicted by them. Kim and Kim [2016a] conducts simultaneous inference of the volatility in the standard factor pricing model in finance. Li et al. [2025b] introduce simultaneous inference of the trend in a time series regression, where the trend is a nonparametric and nonlinear function of an observable random vector.

However, sometime conservative approach like Bonferroni correction would be the choice for simultaneous inference. As an example, the novel dynamics of progressivity poses statistical challenges because the year-by-year estimates of quantile progressivity exhibit several *jumps* (see Figure 4.1), and because the size of our annual time series sample (1978–2022) is limited. The traditional way to construct SCI for an unknown parameter is to rely on some form of *local-smoothing* techniques; see, for example, Wu and Zhao [2007c]. Unfortunately, the local-smoothing would not be a suitable choice here, due to the presence of jumps and the small sample size. While applying Bonferroni adjustment, the problem emerges that inference on the tax progressivity of each single year requires typically high confidence level. To deal with these issues that cannot be revealed a priori, we here employ the high-confidence inference of the tax progressivity. Through this approach, one can avoid the issue of a small sample because the usual local-smoothing for simultaneous inference is no longer required.

The structure of the following content is as follows. Section 4.2 develops simultaneous inference for the quantile regression with high confidence level. Section 4.3 shows an example of U.S. tax progressivity and applying the quantile regression on the income dataset. Section 4.4 concludes. The mathematical proofs are relegated to Appendix.

4.2 Main Results

Notation Denote $\mathbb{1}_{\mathcal{C}}$ as the indicator function. For function $f \in \mathcal{C}^r$, we say f is r -th order continuous derivative. We write $\mathbb{P}(N(0, \Gamma) \in A)$ for the probability that a normally distributed random vector $Z \sim N(0, \Gamma)$ is in the region A . We denote the derivative of a multivariate function as $\nabla g(x)$.

Let

$$g_n(\beta) = \sum_{i=1}^n \psi_\alpha(Y_i - X_i^\top \beta) X_i, \quad g(\beta) = \mathbb{E} g_n(\beta),$$

$$K(\beta) = g_n(\beta) - g(\beta), \quad \Gamma = \text{Cov} \left(\psi_\alpha(Y_1 - X_1^\top \beta) X_1 \right),$$

where $\psi_\alpha(x) = \alpha - \mathbb{1}_{x \leq 0}$. The key results to construct simultaneous confidence interval (SCI) of an unknown parameter function are the following theorem and corollary:

Theorem 4.2.1. *Assume (X_i, Y_i) are i.i.d. with $\mathbb{E}|X_i|^4 < \infty$, $\mathbb{E}|Y_i|^4 < \infty$, $\mathbb{E}X_i X_i^\top > 0$ and the density of e_i given X_i satisfies $0 < m < f_{e_i|X_i}(x) < M < \infty$ for some constant m and M . And $g(\beta) \in \mathcal{C}^2$ with $\sup_\beta |\nabla g(\beta)| + |\nabla^2 g(\beta)| < \infty$. Then,*

$$\sup_{A \in \mathcal{C}} \left| \mathbb{P} \left(\sqrt{n} \nabla g(\beta_0) (\hat{\beta} - \beta_0) \in A \right) - \mathbb{P} (N(0, \Gamma) \in A) \right| \leq C n^{v-1/6},$$

for any $v > 0$, where $\mathcal{C}$ is the class of convex sets in $\mathbb{R}^p$ and C is a constant not depending on n .

Corollary 4.2.1.1. *Let σ_{11}^2 be the first diagonal element of $\tilde{\Gamma}$, where*

$$\tilde{\Gamma} = (\nabla g(\beta_0))^{-1} \Gamma ((\nabla g(\beta_0))^{-1})^\top.$$

Under the same condition of Theorem 4.2.1, for any constant $0 \leq c < \sigma_{11}^2/6$, we have

$$\sup_{0 < t \leq \sqrt{c \log n}} \left| \frac{\mathbb{P}(\sqrt{n}|\hat{\beta}_1 - \beta_{0,1}| \geq t)}{\mathbb{P}(|N(0, \sigma_{11}^2)| \geq t)} - 1 \right| \rightarrow 0.$$

Remark The construction of SCI for the progressivity is a *direct* application of Corollary 4.2.1.1. Our inference is simultaneous under the sup-norm-based statistic in Corollary 4.2.1.1, and is conducted through the SCIs. In contrast to existing inferential methodologies, our SCI-based inference in this study is effective in *suggesting a right function form* for the progressivity. In case a null hypothesis on the tax progressivity, such as its constancy, is somehow rejected, it would be difficult to figure out the *reason for the rejection* if the inference is based on some standard non-simultaneous test statistic and its associated p-value, such as in Härdle and Mammen [1993]. Thus, it would be *non-trivial*, in such a case, to suggest an alternative to the parametric function under the rejected null hypothesis.

4.3 Application: Tax Progressivity

4.3.1 Data

We use the Annual Social and Economic Supplement of the Current Population Survey (CPS-ASEC) from 1978 to 2022 (Flood et al. [2023]) as a baseline. We use income and demographic information from the CPS-ASEC (e.g., age, marital status, relationship to household head, dependent information) and simulate tax liabilities using the National Bureau of Economic Research (NBER) TAXSIM model, which is developed based on actual US fiscal rules.¹

1. We use Version 35, which incorporates federal fiscal laws up to 2023. For details of the program, see Feenberg and Coutts [1993] and <https://taxsim.nber.org/taxsim35/>.

4.3.2 Empirical framework for tax progressivity

Firstly, we introduce the measurement framework for progressivity and our data used in the empirical analysis. We study the dynamics of U.S. tax-and-transfer progressivity across the household income distribution within the class of the parametric effective tax function:

$$T(y) = y - (1 - \lambda) y^{1-\tau}, \quad (4.2)$$

where $T(y)$ is a net tax schedule at income level y , τ is an index of progressivity, and λ governs an average level of taxation in the economy. The large literature on dynamic macroeconomics and public finance has increasingly utilized this effective tax function to study the implications of progressive taxation because of its merits: the two-parameter function offers a reasonable fit to the empirical relationship between market income and disposable income, and tractability in modeling progressive taxation in quantitative models (e.g., Feldstein [1969]; Benabou [2002]; Heathcote et al. [2017]).

In an empirical implementation, a popular way is to estimate $1 - \tau$, which is the residual elasticity of post-government (disposable) income $y - T(y)$ with respect to pre-government (market) income y .² Estimation of $1 - \tau$ is conducted through the regression:

$$y_{it} = \beta_0 + \beta_1 x_{it} + e_{it}, \quad i = 1, \dots, n_t, \quad t = 1, \dots, T, \quad (4.3)$$

where i is a cross-sectional unit and t is a time index. The constant regression coefficients β_0 and β_1 correspond to $\log(1 - \lambda)$ and $1 - \tau$, respectively. Assuming that the regression coefficients are time-invariant, researchers have obtained least-squares estimates for a value of $1 - \tau$. Because the least-squares method estimates the conditional mean, the resulting estimate represents the *mean* residual elasticity, which provides a representative progressivity

2. Following the terminology in Heathcote et al. [2017], we use “disposable income” interchangeably with “post-government income” and “market income” with “pre-government income,” respectively.

in a limited case where the mean progressivity is required to be similar across the distribution.

We generalize this widely used approach in the current literature by allowing the regression coefficients to *vary* over time *and* across the income distribution. Formally, we measure the progressivity through the generalized regression framework:

$$y_{it} = \beta_0(\alpha, t) + \beta_1(\alpha, t) x_{it} + e_{it}, \quad i = 1, \dots, n_t, \quad t = 1, \dots, T, \quad (4.4)$$

where α is a quantile of the conditioned income distribution. In equation (4.4), regression coefficients $\beta_0(\alpha, t)$ and $\beta_1(\alpha, t)$ *vary* by *both* time and income percentile. A key advantage of our generalized framework is that it allows us to trace varying curvature across income quantiles, thereby addressing the well-known limitations of the least-squares approach in estimating the tax function.³ Moreover, this flexible time-varying quantile framework is robust to changes in the structure of tax and transfer system, which enables us to capture the potential non-linear dynamics across the income distribution.

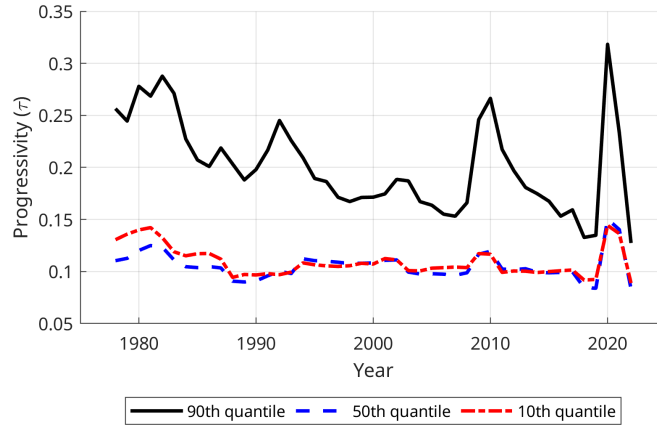


Figure 4.1: Yearly Progressivity Estimates by Income Quantiles

Figure 4.1 shows the U.S. yearly progressivity parameter estimates from 1978 to 2022 at three income percentiles—the 10th, 50th, and 90th. As evident in the figure, there is

3. The least-squares estimate of the tax function is known to misfit the *bottom* of the income distribution and to underestimate progressivity at low-income levels (Heathcote et al. [2017]).

a *substantial* amount of time variation in the progressivity estimates for all three income quantiles. Interestingly, the extent of time variation in the estimated progressivity appears to be *income-dependent*: the higher the income percentile, the greater the time variation in the estimated quantile progressivity, especially largest at the 90th percentile household income.

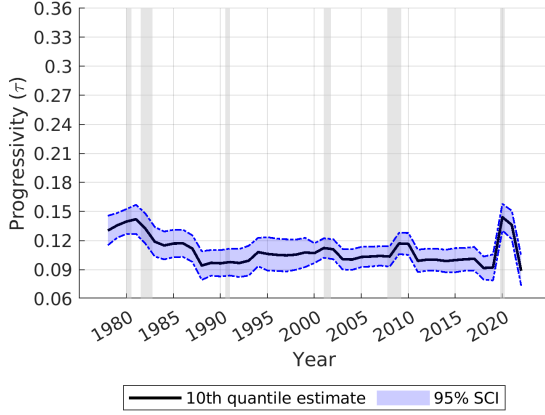
4.3.3 Inference for dynamic quantile progressivity

Building on the theoretical results, we construct simultaneous confidence intervals (SCIs) for the U.S. tax progressivity parameter at quantiles of the conditional income distribution. Specifically, we examine three different income quantiles—the 10th, 50th, and 90th. Figure 4.2 presents the results, overlaid with NBER recession periods to highlight the potential influence of U.S. business cycles on tax progressivity.

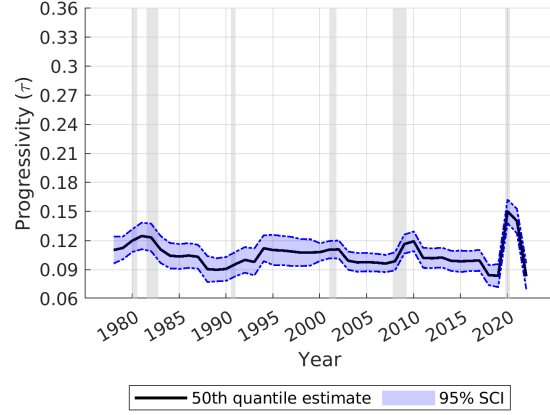
As we can see from Figure 4.2, the progressivity generally *rises during recessions* (i.e. the shaded vertical stripes in the figure). The main advantage of SCI is that the interval can be used to test any hypothesis of a function form for the progressivity. For example, the null hypothesis of a constant progressivity is rejected if the constructed 95% SCI fails to contain any horizontal line *entirely* in it. For this reason, the hypothesis of a time-invariant progressivity is *rejected* at 5% level for all three income quantiles. In Figure C.1 in Appendix C.1, we also display the simultaneous confidence intervals at the first and third quartiles and show that the cross-sectional pattern of volatility is consistent. More importantly, the tight SCIs allow us to statistically validate that most of the time variation stems from progressivity at the 90th conditional quantile, given its notable fluctuations over time.

Methodologically, since the inferential methodology developed in Section 4.2 is free of local-smoothing, this new inference methodology is less affected by the small-sample issue. Hence one can further identify the *exact* sub-periods where the assumption of constancy is rejected, simply by *locating where the constancy is violated* (i.e. where the SCI fails to

(a) 10th percentile



(b) 50th percentile



(c) 90th percentile

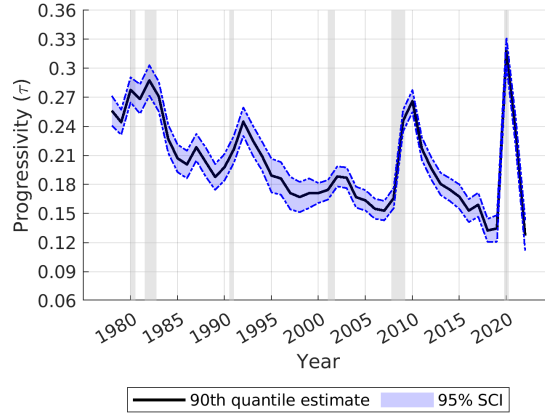


Figure 4.2: Progressivity Estimates and 95% Simultaneous Confidence Intervals

Notes: Panels (a), (b), and (c) show the U.S. tax progressivity estimates at each income percentile (black solid curve) and 95% simultaneous confidence bands obtained using the simultaneous inference method (blue area). Grey-shaded regions indicate NBER recessions.

contain the horizontal line). At all income percentiles, a declining trend in the 1980s and all recessions stand out as the periods of a *notably time-varying* progressivity, so that the null of a constant progressivity is rejected mainly due to these episodes. In the 1980s, under the Reagan administration, the U.S. government implemented the across-the-board tax cuts, potentially contributing to a decline in the progressivity.⁴ A key methodological advancement

4. Borella et al. [2023] relate various U.S. tax reforms and the time variation in the tax progressivity based on year-by-year mean progressivity point estimates alone, without statistical validation.

over the existing studies is our use of the newly developed inferential methodology in Section ??, which allows one to statistically determine the specific episodes and quantile of the conditional distribution that drive notable changes in tax progressivity.

As for the late 2000s and the early 2020s, the period overlaps with the Great Recession and the COVID-19 pandemic during which the U.S. government introduced substantial relief packages. Determining whether or not the observed time variation in these deep recessions induces structural changes akin to those of the 1980s or is merely a consequence of the strong counter-cyclicalities is an important question. For instance, if the changes around 2020 are a temporary result of the strong counter-cyclicalities, one would consider that a stationary process would be suitable for modeling the time variation around the COVID-19 pandemic. In contrast, if progressivity experienced any permanent changes around 2020, then there would be fundamentally different welfare and aggregate implications. Our methodology would be essential to answer the question by suggesting an appropriate function form with statistical validation based on data.

4.4 Conclusion

In this chapter, we address the statistical challenge of constructing high-confidence-level inference in quantile regression. Classical inference methods, often relying on central limit theorems and local smoothing, become unreliable in such settings—especially when confidence levels approach one. To this end, we develop a moderate deviation theory for quantile regression estimators and establish a framework for simultaneous confidence intervals that remain valid in the extreme tails of the distribution. Our method allows for robust inference without the need for local smoothing, and it provides a powerful tool to detect and characterize time variation in dynamic systems.

We apply our methodology to study the dynamics of U.S. tax progressivity from 1978 to 2022. Using a quantile regression framework, we estimate year-by-year progressivity

across income percentiles and construct simultaneous confidence intervals to assess the time variation statistically. We reject the null hypothesis of constant progressivity, especially during major recessions and policy reform periods such as the 1980s. These results highlight that U.S. tax progressivity is not only time-varying but also responsive to macroeconomic conditions and structural changes, and that our method can reliably identify such dynamics with statistical rigor.

REFERENCES

- Daron Acemoglu and David Autor. *Skills, Tasks and Technologies: Implications for Employment and Earnings*, volume 4, pages 1043–1171. Elsevier, 2011.
- Christos Alexopoulos and David Goldsman. To batch or not to batch? *ACM Transactions on Modeling and Computer Simulation*, 14(1):76–114, jan 2004. ISSN 1049-3301. doi:10.1145/974734.974738. URL <https://doi.org/10.1145/974734.974738>.
- Christos Alexopoulos, David Goldsman, Anup C Mokashi, Kai-Wen Tien, and James R Wilson. Sequest: A sequential procedure for estimating quantiles in steady-state simulations. *Operations Research*, 67(4):1162–1183, 2019.
- Fernando Alvarez, Andrew Atkeson, and Patrick J. Kehoe. Time-varying risk, interest rates, and exchange rates in general equilibrium. *Review of Economic Studies*, 76:851–878, 2009.
- NN Amosova. Local limit theorems for probabilities of moderate deviations. *Lithuanian mathematical transactions of the Academy of Sciences of the Lithuanian SSR*, 14(3):401–407, 1974.
- Jiří Anděl, Ivan Netuka, and Karel Zvára. On threshold autoregressive processes. *Kybernetika*, 20(2):89–106, 1984.
- Tomohiro Ando, Matthew Greenwood-Nimmo, and Yongcheol Shin. Quantile connectedness: modeling tail behavior in the topology of financial networks. *Management Science*, 68(4):2401–2431, 2022.
- G. Aneiros-Pérez and J. M. Vilar-Fernández. Local polynomial estimation in partial linear regression models under dependence. *Computational Statistics & Data Analysis*, 52(5):2757–2777, January 2008.
- Andrew Ang and Dennis Kristensen. Testing conditional factor models. *Journal of Financial Economics*, 106:132–156, 2012.
- Florin Avram and Murad S Taqqu. Noncentral limit theorems and appell polynomials. *The Annals of Probability*, 15(2):767–775, 1987.
- Richard T. Baillie and Kun Ho Kim. Was it risk? or was it fundamentals? explaining excess currency returns with kernel smoothed regressions. *Journal of Empirical Finance*, 34:99–111, 2015a.
- Richard T. Baillie and Kun Ho Kim. Was it risk? or was it fundamentals? explaining excess currency returns with kernel smoothed regressions. *Journal of Empirical Finance*, 34:99–111, 2015b.
- Richard T. Baillie, Francis X. Diebold, George Kapetanios, and Kun Ho Kim. A new test for market efficiency and uncovered interest parity. *Journal of International Money and Finance*, 130(102765), 2023.

- Ozan Bakış, Barış Kaymak, and Markus Poschke. Transitional dynamics and the optimal progressivity of income redistribution. *Review of Economic Dynamics*, 18:679–693, 2015. ISSN 1094-2025.
- Fabio Bellini, Francesco Cesarone, Christian Colombo, and Fabio Tardella. Risk parity with expectiles. *European Journal of Operational Research*, 291(3):1149–1163, 2021.
- Denis Belomestny and Volker Krätschmer. Optimal stopping under probability distortions. *Mathematics of Operations Research*, 42(3):806–833, 2017.
- Roland Benabou. Tax and education policy in a heterogeneous-agent economy: What levels of redistribution maximize growth and efficiency? *Econometrica*, 70:481–517, 3 2002.
- István Berkes, Weidong Liu, and Wei Biao Wu. Komlós-Major-Tusnády approximation under dependence. *The Annals of Probability*, 42(2):794–817, 2014.
- Anmol Bhandari, David Evans, Mikhail Golosov, and Thomas J Sargent. Inequality, business cycles, and monetary-fiscal policy. *Econometrica*, 89:2559–2599, 11 2021. ISSN 0012-9682. doi:<https://doi.org/10.3982/ECTA16414>. URL <https://doi.org/10.3982/ECTA16414>.
- Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.
- John F.O. Bilson. The speculative efficiency hypothesis. *The Journal of Business*, 54(3): 435–451, 1981.
- Nicholas H Bingham, Charles M Goldie, and Jef L Teugels. *Regular variation*. Number 27. Cambridge university press, 1989.
- Corina Boar and Virgiliu Midrigan. Efficient redistribution. *Journal of Monetary Economics*, 131:78–91, 2022. ISSN 0304-3932.
- Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.
- Soham Bonnerjee, Johannes Moritz Jirak, Wei Biao Wu, and Han Xiao. Gaussian approximation for high-dimensional linear forms of stationary processes. *Manuscript*, 2024.
- Margherita Borella, Mariacristina De Nardi, Michael Pak, Nicolo Russo, and Fang Yang. The importance of modeling income taxes over time: U.s. reforms and outcomes. *Journal of the European Economic Association*, 21:2237–2286, 12 2023.
- George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- Richard C. Bradley. Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions. *Probability Surveys*, 2:107 – 144, 2005a. doi:10.1214/154957805100000104. URL <https://doi.org/10.1214/154957805100000104>.

- Richard C. Bradley. Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions. *Probability Surveys*, 2(none):107 – 144, 2005b. doi:10.1214/154957805100000104. URL <https://doi.org/10.1214/154957805100000104>.
- Donald L. Burkholder. Sharp inequalities for martingales and stochastic integrals. In *Colloque Paul Lévy sur les processus stochastiques*, number 157-158 in Astérisque. 1988.
- Mine Çağlar. A long-range dependent workload model for packet data traffic. *Mathematics of Operations Research*, 29(1):92–105, 2004.
- Li Cai and Suojin Wang. Simultaneous confidence bands and global inferences for extended partially linear single-index models. *Applied Mathematical Modelling*, 113:30–43, January 2023.
- Ming Chen and Qiongxia Song. Simultaneous inference of the mean of functional time series. *Electronic Journal of Statistics*, 9(2):1779 – 1798, 2015.
- Zhi Chen, Anil Gaba, Ilia Tsetlin, and Robert L. Winkler. Evaluating quantile forecasts in the m5 uncertainty competition. *International Journal of Forecasting*, 38(4):1531–1545, 2022. ISSN 0169-2070. doi:<https://doi.org/10.1016/j.ijforecast.2022.03.004>. URL <https://www.sciencedirect.com/science/article/pii/S0169207022000449>. Special Issue: M5 competition.
- Victor Chernozhuokov, Denis Chetverikov, Kengo Kato, and Yuta Koike. Improved central limit theorem and bootstrap approximations in high dimensions. *The Annals of Statistics*, 50(5):2562–2586, 2022.
- Peter F Christoffersen and Francis X Diebold. Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science*, 52(8):1273–1287, 2006.
- Jean-François Coeurjolly and Hedi Kortas. Expectiles for subordinated Gaussian processes with applications. *Electronic Journal of Statistics*, 6:303 – 322, 2012. doi:10.1214/12-EJS674. URL <https://doi.org/10.1214/12-EJS674>.
- Abdelaati Daouia, Stéphane Girard, and Gilles Stupfler. Tail expectile process and risk assessment. *Bernoulli*, 26(1):531–556, 2020.
- Richard A Davis and Tailen Hsing. Point process and partial sum convergence for weakly dependent random variables with infinite variance. *The Annals of Probability*, 23(2):879–917, 1995.
- James Dawber, Nicola Salvati, Enrico Fabrizi, and Nikos Tzavidis. Expectile regression for multi-category outcomes with application to small area estimation of labour force participation. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 185 (Supplement_2):S590–S619, 2022.

- Alexandre Barbosa De Lima and José Roberto de Almeida Amazonas. *Internet Teletraffic Modeling and Estimation*. River Publishers, 2013.
- Jérôme Dedecker and Clémentine Prieur. New dependence coefficients. examples and applications to statistics. *Probability Theory and Related Fields*, 132:203–236, 2005.
- Erick Delage, Daniel Kuhn, and Wolfram Wiesemann. “dice”-sion-making under uncertainty: When can a random decision reduce risk? *Management Science*, 65(7):3282–3301, 2019.
- Gerhard Derflinger, Wolfgang Hörmann, and Josef Leydold. Random variate generation by numerical inversion when only the density is known. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 20(4):1–25, 2010.
- Kemal Dincer Dinceç, Christos Alexopoulos, David Goldsman, Athanasios Lolos, and James R. Wilson. *Geometric-Moment Contraction of $G/G/1$ Waiting Times*, pages 111–130. Springer International Publishing, Cham, 2022. ISBN 978-3-031-10193-9. doi:10.1007/978-3-031-10193-9_6. URL https://doi.org/10.1007/978-3-031-10193-9_6.
- Paul Doukhan. *Mixing: properties and examples*, volume 85. Springer Science & Business Media, 2012.
- Paul Doukhan and Sana Louhichi. A new weak dependence condition and applications to moment inequalities. *Stochastic processes and their applications*, 84(2):313–342, 1999.
- Paul Embrechts, Haiyan Liu, and Ruodu Wang. Quantile-based risk sharing. *Operations Research*, 66(4):936–949, 2018.
- Paul Embrechts, Alexander Schied, and Ruodu Wang. Robustness in the optimization of risk measures. *Operations Research*, 70(1):95–110, 2022.
- Susanne Emmer, Marie Kratz, and Dirk Tasche. What is the best risk measure in practice? a comparison of standard measures. *Journal of Risk*, (2):31–60, 2015.
- Charles Engel. The forward discount anomaly and the risk premium: A survey of recent evidence. *Journal of Empirical Finance*, 3,:123–192, 1996.
- Joseph Engelberg, David R. McLean, and Jeffrey Pontiff. Anomalies and news. *Journal of Finance*, 73(5):1971–2001, 2018.
- Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50, no. 4:987–1007, 1982.
- Robert F. Engle, C. W. J. Granger, John Rice, and Andrew Weiss. Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association*, 81(394):310–320, 1986a.

- Robert F. Engle, C. W. J. Granger, John Rice, and Andrew Weiss. Semiparametric Estimates of the Relation Between Weather and Electricity Sales. *Journal of the American Statistical Association*, 81(394):310–320, 1986b.
- R. L. Eubank and P. L. Speckman. Confidence Bands in Nonparametric Regression. *Journal of the American Statistical Association*, 88(424):1287–1301, 1993.
- Eugene Fama. Forward and spot exchange rates. *Journal of Monetary Economics*, 14: 319–338, 1984.
- J. Fan and T. Huang. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli*, 11:1031–1057, 2005a.
- J. Fan and W. Zhang. Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scandinavian Journal of Statistics*, 27:715–731, 2000.
- Jianqing Fan and I. Gijbels. *Local Polynomial Modelling and its Applications*. Springer New York, NY, 1 edition, 1996.
- Jianqing Fan and Tao Huang. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli*, 11(6):1031–1057, December 2005b. Publisher: Bernoulli Society for Mathematical Statistics and Probability.
- Yanqin Fan and Qi Li. Root-n-consistent estimation of partially linear time series models. *Journal of Nonparametric Statistics*, 11(1-3):251–269, 1999.
- Yanqin Fan and Qi Li. A Kernel-Based Method for Estimating Additive Partially Linear Models. *Statistica Sinica*, 13(3):739–762, 2003. Publisher: Institute of Statistical Science, Academia Sinica.
- Daniel Feenberg and Elisabeth Coutts. An introduction to the taxsim model. *Journal of Policy Analysis and management*, 12:189–194, 1993.
- Martin S Feldstein. The effects of taxation on risk taking. *Journal of Political Economy*, 77: 755–764, 9 1969.
- T. S. Ferguson. *A Course in large sample theory*. Chapman and Hall, CRC, 1996.
- Axelle Ferriere and Gaston Navarro. The heterogeneous effects of government spending: It’s all about taxes. *The Review of Economic Studies*, page rdae032, 4 2024. ISSN 0034-6527. doi:10.1093/restud/rdae032. URL <https://doi.org/10.1093/restud/rdae032>.
- Johannes Fleck, Jonathan Heathcote, Kjetil Storesletten, and Giovanni L. Violante. Tax and transfer progressivity at the us state level. *Manuscript*, 2021.
- Sarah Flood, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Megan Schouweiler, and Michael Westberry. Ipums cps: Version 11.0 [dataset], 2023.

- Jürgen Franke, Wolfgang Karl Härdle, and Christian M Hafner. *Statistics of financial markets*, volume 2. Springer, 2004.
- Kenneath A. Froot and Richard H. Thaler. Anomalies: Foreign exchange. *Journal of Economic Perspectives*, 4(3):179–192, 1990a.
- Kenneth A. Froot and Richard H. Thaler. Anomalies: Foreign exchange. *Journal of Economic Perspectives*, 4:179–192, 1990b.
- Jiti Gao, Shengyan Hong, and Hua Liang. Convergence rates of a class of estimators in partly linear models. *Acta Mathematica Sinica*, 38, 01 1995.
- Rosella Giacometti, Gabriele Torri, and Sandra Paterlini. Tail risks in large portfolio selection: penalized quantile and expectile minimum deviation models. *Quantitative Finance*, 21(2):243–261, 2021.
- Georgios B Giannakis and Erchin Serpedin. A bibliography on nonlinear system identification. *Signal Processing*, 81(3):533–580, 2001.
- David Goldsman and Lee Schruben. Asymptotic properties of some confidence interval estimators for simulation output. *Management Science*, 30(10):1217–1225, 1984.
- Peter J. Green and Brian S. Yandell. Semi-parametric Generalized Linear Models. In Robert Gilchrist, Brian Francis, and Joe Whittaker, editors, *Generalized Linear Models*, Lecture Notes in Statistics, pages 44–55. Springer US, 1985.
- Yael Grushka-Cockayne, Kenneth C Lichtendahl Jr, Victor Richmond R Jose, and Robert L Winkler. Quantile evaluation, sensitivity to bracketing, and sharing business payoffs. *Operations Research*, 65(3):712–728, 2017.
- Nezih Guner, Remzi Kaygusuz, and Gustavo Ventura. Income taxation of u.s. households: Facts and parametric estimates. *Review of Economic Dynamics*, 17(4):559–581, 2014. ISSN 1094-2025. doi:<https://doi.org/10.1016/j.red.2014.01.003>. URL <https://www.sciencedirect.com/science/article/pii/S1094202514000040>.
- Mengmeng Guo and Wolfgang Karl Härdle. Simultaneous confidence bands for expectile functions. *AStA Advances in Statistical Analysis*, 96:517–541, 2012.
- James D. Hamilton. Why you should never use the hodrick-prescott filter. *Review of Economics and Statistics*, 100:831–843, 2018.
- Scott A. Hamilton and Young K. Truong. Local Linear Estimation in Partly Linear Models. *Journal of Multivariate Analysis*, 60(1):1–19, January 1997.
- W. Härdle and E. Mammen. Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, 21(4):1926–1947, 1993.
- W Härdle, H. Liang, and J. Gao. *Partially linear models*. Heidelberg, Physica-Verlag, 2000.

- Wolfgang Härdle, Hua Liang, and Jiti Gao. *Partially Linear Models*. Contributions to Statistics. Physica-Verlag HD, Heidelberg, 2000.
- Wolfgang K. Härdle. Asymptotic maximal deviation of M-smoothers. *Journal of Multivariate Analysis*, 29(2):163–179, May 1989a.
- Wolfgang K. Härdle. Asymptotic maximal deviation of M-smoothers. *Journal of Multivariate Analysis*, 29(2):163–179, May 1989b.
- Wolfgang K. Härdle and Song Song. Confidence bands in quantile regression. *Econometric Theory*, 26(4):1180–1200, 2010.
- Wolfgang Karl Härdle and Enno Mammen. Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, 21(4):1926–1947, 1993.
- Campbell Harvey. Time-varying conditional covariances in tests of asset pricing models. *Journal of Financial Economics*, 24:289–317, 1989.
- Jonathan Heathcote, Fabrizio Perri, and Giovanni L Violante. Unequal we stand: An empirical analysis of economic inequality in the united states, 1967–2006. *Review of Economic Dynamics*, 13:15–51, 2010.
- Jonathan Heathcote, Kjetil Storesletten, and Giovanni L. Violante. Optimal tax progressivity: An analytical framework. *Quarterly Journal of Economics*, 132:1693–1754, 2017.
- Jonathan Heathcote, Kjetil Storesletten, and Giovanni L. Violante. Optimal progressivity with age-dependent taxation. *Journal of Public Economics*, 189, 2020a.
- Jonathan Heathcote, Kjetil Storesletten, and Giovanni L. Violante. How should tax progressivity respond to rising income inequality? *Journal of the European Economic Association*, 18:2715–2754, 2020b.
- Nancy E. Heckman. Spline smoothing in a partly linear model. *Journal of the Royal Statistical Society: Series B (Methodological)*, 48(2):244–248, 1986.
- Bruce M. Hill. A Simple General Approach to Inference About the Tail of a Distribution. *The Annals of Statistics*, 3(5):1163 – 1174, 1975. doi:10.1214/aos/1176343247. URL <https://doi.org/10.1214/aos/1176343247>.
- Hwai-Chung Ho and Fang-I Liu. Estimation of short-and long-term var for long-memory stochastic volatility models. *Handbook of Quantitative Finance and Risk Management*, pages 1409–1415, 2010.
- Robert J. Hodrick. Risk, uncertainty and exchange rates. *Journal of Monetary Economics*, 23:433–459, 1989.
- Robert J. Hodrick and Edward C. Prescott. Postwar u.s. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, 29:1–16, 1997.

- Hajo Holzmann and Bernhard Klar. Expectile asymptotics. *Electronic Journal of Statistics*, 10(2):2355–2371, 2016.
- L Jeff Hong. Estimating quantile sensitivities. *Operations research*, 57(1):118–130, 2009.
- Xueyan Hu, Wei Liu, and Hua Huo. An intelligent network traffic prediction method based on butterworth filter and cnn-lstm. *Computer Networks*, 240:110172, 2024. ISSN 1389-1286. doi:<https://doi.org/10.1016/j.comnet.2024.110172>. URL <https://www.sciencedirect.com/science/article/pii/S1389128624000045>.
- Peter J Huber. Robust regression: asymptotics, conjectures and monte carlo. *The annals of statistics*, pages 799–821, 1973.
- Peter J Huber. Robust statistics. In *International encyclopedia of statistical science*, pages 1248–1251. Springer, 2011.
- Ravi Jagannathan and Zhenyu Wang. The conditional capm and the cross-section of expected returns. *Journal of Finance*, 51(1):3–53, 1996.
- Youngsoo Jang, Takeki Sunakawa, and Minchul Yum. Tax-and-transfer progressivity and business cycles. *Quantitative Economics*, 14:1367–1400, 11 2023. ISSN 1759-7323. doi:<https://doi.org/10.3982/QE1568>. URL <https://doi.org/10.3982/QE1568>.
- Narasimhan Jegadeesh and Sheridan Titman. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48(1):65–91, 1993.
- Gordon J. Johnston. Probabilities of maximal deviations for nonparametric regression function estimates uncertainty and exchange rates. *Journal of Multivariate Analysis*, 12(3):402–414, 1982a.
- Gordon J Johnston. Probabilities of maximal deviations for nonparametric regression function estimates. *Journal of Multivariate Analysis*, 12(3):402–414, September 1982b.
- Gordon J Johnston. Probabilities of maximal deviations for nonparametric regression function estimates. *Journal of Multivariate Analysis*, 12(3):402–414, September 1982c.
- M Chris Jones. Expectiles and m-quantiles are quantiles. *Statistics & Probability Letters*, 20(2):149–153, 1994.
- Victor Richmond R Jose and Robert L Winkler. Evaluating quantile assessments. *Operations research*, 57(5):1287–1297, 2009.
- Jinyong Kim, Kun Ho Kim, and Taejin Kim. New insights on the forward premium regression through a portfolio-based approach. *CIREQ Working Paper Series*, 2023.
- Kun Ho Kim. Inference of the trend in a partially linear model with locally stationary regressors. *Econometric Reviews*, 35:1194–1220, 2016a.

- Kun Ho Kim. Inference of the trend in a partially linear model with locally stationary regressors. *Econometric Reviews*, 35(7):1194–1220, 2016b.
- Kun Ho Kim and Taejin Kim. Capital asset pricing model: A time-varying volatility approach. *Journal of Empirical Finance*, 37:268–281, 2016a.
- Kun Ho Kim and Taejin Kim. Capital asset pricing model: a time-varying volatility approach. *Journal of Empirical Finance*, 37:268–281, 2016b.
- Kun Ho Kim, Shih-Kang Chao, and Wolfgang K. Härdle. Simultaneous inference of the partially linear model with a multivariate unknown function. *Journal of Statistical Planning and Inference*, 213:93–105, 09 2021a.
- Kun Ho Kim, Shih-Kang Chao, and Wolfgang K. Härdle. Simultaneous inference of the partially linear model with a multivariate unknown function. *Journal of Statistical Planning and Inference*, 213:93–105, July 2021b.
- George Knafl, Jerome Sacks, and Donald Ylvisaker. Confidence Bands for Regression Functions. *Journal of the American Statistical Association*, 80(391):683–691, 1985. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].
- Roger Koenker. *Quantile regression*, volume 38. Cambridge university press, 2005.
- Roger Koenker and Gilbert Bassett. Regression quantiles. *Econometrica*, 50:43–61, 1978.
- Johannes Konig. Bias in tax progressivity estimates. *National Tax Journal*, 76:267–289, 5 2023.
- Gizem Kosar and Robert A Moffitt. Trends in cumulative marginal tax rates facing low-income families, 1997–2007. *Tax Policy and the Economy*, 31:43–70, 1 2017. ISSN 0892-8649. doi:10.1086/691083. doi: 10.1086/691083.
- Volker Krättschmer and Henryk Zähle. Statistical inference for expectile-based risk measures. *Scandinavian Journal of Statistics*, 44(2):425–454, 2017.
- Chung-Ming Kuan, Jin-Huei Yeh, and Yu-Chin Hsu. Assessing value at risk with care, the conditional autoregressive expectile models. *Journal of Econometrics*, 150(2):261–270, 2009.
- Soumendra Nath Lahiri. *Resampling methods for dependent data*. Springer Science & Business Media, 2003.
- Averill M Law. *Simulation modeling and analysis Fifth edition*. McGraw-Hill New York, 2013.
- Lei Lei, Christos Alexopoulos, Yijie Peng, and James R Wilson. Estimating confidence intervals and regions for quantiles by monte carlo simulation. *Available at SSRN 3959456*, 2021.

- Jonathan Lewellen and Stefan Nagel. The conditional capm does not explain asset pricing anomalies. *Journal of Financial Economics*, 82(1):289–314, 2006.
- Degui Li, Weidong Liu, Qiying Wang, and Wei Biao Wu. Simultaneous confidence bands in nonlinear regression models with nonstationarity. *Statistica Sinica*, 27(3):1385–1400, 2017.
- Jiaqi Li, Likai Chen, Kun Ho Kim, and Tianwei Zhou. Simultaneous inference of a partially linear model in time series. *Journal of Time Series Analysis*, 2025a.
- Jiaqi Li, Likai Chen, Kun Ho Kim, and Tianwei Zhou. Simultaneous inference of a partially linear model in time series. *Journal of Time Series Analysis*, 2025b.
- Wai Keung Li, Shiqing Ling, and Michael McAleer. Recent theoretical results for time series models with garch errors. *Journal of Economic Surveys*, 16(3):245–269, 2002.
- Xiaocheng Li, Huaiyang Zhong, and Margaret L Brandeau. Quantile markov decision processes. *Operations research*, 70(3):1428–1447, 2022.
- Hua Liang, Wolfgang Karl Härdle, and Axel Werwatz. Asymptotic properties of the non-parametric part in partial linear heteroscedastic regression models. *RePEc*, 01 1997a.
- Hua Liang, Wolfgang Karl Härdle, and Axel Werwatz. Asymptotic properties of the non-parametric part in partial linear heteroscedastic regression models. *RePEc*, 01 1997b.
- Hua Liang, Yongsong Qin, Xinyu Zhang, and David Ruppert. Empirical Likelihood-Based Inferences for Generalized Partially Linear Models. *Scandinavian Journal of Statistics*, 36(3):433–443, 2009.
- Wu Liming. Moderate deviations of dependent random variables related to clt. *The Annals of Probability*, pages 420–445, 1995.
- Scott C Linn and Nicholas SP Tay. Complexity and the character of stock returns: Empirical evidence and a model of asset prices based on complex investor learning. *Management Science*, 53(7):1165–1180, 2007.
- Fangda Liu and Ruodu Wang. A theory for measures of tail risk. *Mathematics of Operations Research*, 46(3):1109–1128, 2021.
- Weidong Liu and Wei Biao Wu. Simultaneous nonparametric inference of time series. *The Annals of Statistics*, 38(4):2388–2421, 2010a.
- Weidong Liu and Wei Biao Wu. Simultaneous nonparametric inference of time series. *The Annals of Statistics*, 38(4):2388–2421, June 2010b.
- Robert E Lucas. *Models of Business Cycles*. Oxford: Basil Blackwell, 1987.
- Hanno Lustig, Nikolai Roussanov, and Adrian Verdelhan. Countercyclical currency risk premia. *Journal of Financial Economics*, 111:527–553, 2014.

- Leandro De Magalhaes, Enric Martorell, and Raul Santaaulalia-Llopi. Transfer progressivity and development. *Manuscript*, 2022.
- Nelson C. Mark. Exchange rates and fundamentals: Evidence on long-horizon prediction. *American Economic Review*, 85:201–218, 1995.
- Alisdair McKay and Ricardo Reis. The role of automatic stabilizers in the u.s. business cycle. *Econometrica*, 84(1):141–194, 2016. doi:<https://doi.org/10.3982/ECTA11574>. URL <https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA11574>.
- Alisdair McKay and Ricardo Reis. Optimal automatic stabilizers. *The Review of Economic Studies*, 88:2375–2406, 10 2021. ISSN 0034-6527. doi:10.1093/restud/rdaa038. URL [http s://doi.org/10.1093/restud/rdaa038](https://doi.org/10.1093/restud/rdaa038).
- Ian Melbourne. Large and moderate deviations for slowly mixing dynamical systems. *Proceedings of the American Mathematical Society*, 137(5):1735–1741, 2009.
- Francisco D Munoz, Benjamin F Hobbs, and J-P Watson. New bounding and decomposition approaches for milp investment problems: Multi-area transmission and generation planning under policy constraints. *European Journal of Operational Research*, 248(3):888–898, 2016.
- Fedor Nazarov. On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a gaussian measure. In *Geometric Aspects of Functional Analysis: Israel Seminar 2001-2002*, pages 169–187. Springer, 2003.
- Whitney K. Newey and James L. Powell. Asymmetric least squares estimation and testing. *Econometrica*, 55(4):819–847, 1987. ISSN 00129682, 14680262. URL <http://www.jstor.org/stable/1911031>.
- Des F Nicholls and Barry G Quinn. *Random Coefficient Autoregressive Models: An Introduction*. Springer Science & Business Media, 2012.
- Jeh-Nan Pan and Su-Tsu Chen. Monitoring long-memory air quality data using arfima model. *Environmetrics*, 19(2):209–219, 2008.
- Nataliya D Pankratova, Nataliia G Zrazhevskaya, et al. Method of dynamic var and cvar risk measures forecasting for long range dependent time series on the base of the heteroscedastic model. *Intelligent Control and Automation*, 8(02):126–138, 2017.
- Collin Philipps. An expectile weak law of large numbers. *Available at SSRN 4302890*, 2022a.
- Collin Philipps. A uniform ergodic theorem for expectiles. *Available at SSRN 4302891*, 2022b.
- Collin Philipps. Quasi-maximum likelihood estimation for conditional expectiles. *Available at SSRN 4302896*, 2023.

- Thomas Piketty and Emmanuel Saez. How progressive is the u.s. federal tax system? a historical and international perspective. *Journal of Economic Perspectives*, 21:3–24, 3 2007.
- Dimitris N Politis, Joseph P Romano, and Michael Wolf. *Subsampling*. Springer Science & Business Media, 2012.
- M. B. Priestley. *Non-linear and Non-stationary Time Series Analysis*. Academic Pr, 1988.
- M.B. Priestley. *Non-linear and Non-stationary Time Series Analysis*. Probability and mathematical statistics. Academic Press. ISBN 9780125649100.
- Xincheng Qiu and Nicolo Russo. Income taxation: A cross-country comparison. *Manuscript*, 2022.
- Martin Raič. A multivariate berry-esseen theorem with explicit constants. *Bernoulli*, 25: 2824–2853, 09 2019. doi:10.3150/18-BEJ1072.
- Rui Ren, Meng-Jou Lu, Yingxing Li, and Wolfgang Karl Härdle. Financial Risk Meter FRM based on expectiles. *Journal of Multivariate Analysis*, 189:104881, 2022.
- Sidney Resnick and Gennady Samorodnitsky. A heavy traffic approximation for workload processes with heavy tailed service requirements. *Management Science*, 46(9):1236–1248, 2000.
- John Rice. Convergence rates for partially splined models. *Statistics & Probability Letters*, 4(4):203–208, June 1986.
- Emmanuel Rio. Moment inequalities for sums of dependent random variables under projective conditions. *Journal of Theoretical Probability*, 22:146–163, 01 2009.
- P. M. Robinson. Root-n-consistent semiparametric regression. *Econometrica*, 56(4):931–954, 1988a.
- Peter M. Robinson. Root-n-consistent semiparametric regression. *Econometrica*, 56:931–954, 1988b.
- M Rosenblatt. Stationary processes as shifts of functions of independent random variables. *Journal of Mathematics and Mechanics*, pages 665–681, 1959.
- Murray Rosenblatt. Remarks on some nonparametric estimates of a density function. *The annals of mathematical statistics*, pages 832–837, 1956.
- Murray Rosenblatt. Functions of markov processes. *Markov Processes. Structure and Asymptotic Behavior*, pages 57–78, 1971.
- Herman Rubin and Jayaram Sethuraman. Probabilities of moderate deviations. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 325–346, 1965.

- Emmanuel Saez and Gabriel Zucman. The rise of income and wealth inequality in america: Evidence from distributional macroeconomic accounts. *Journal of Economic Perspectives*, 34:3–26, 2020.
- Anton Schick. Root-n consistent estimation in partly linear regression models. *Statistics & Probability Letters*, 28(4):353–358, August 1996.
- Bruce Schmeiser. Batch size effects in the analysis of simulation output. *Operations Research*, 30(3):556–568, 1982.
- Nadhemi Selmi and Nejib Hachicha. Long range dependency and forecasting of housing price index and mortgage market rate: evidence of subprime crisis. *Management Science Letters*, 5(5):419–430, 2015.
- Thomas A. Severini and Joan G. Staniswalis. Quasi-likelihood Estimation in Semiparametric Models. *Journal of the American Statistical Association*, 89(426):501–511, 1994. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].
- Xiaofeng Shao and Wei Biao Wu. Asymptotic spectral theory for nonlinear time series. *The Annals of Statistics*, 35(4):1773–1801, 2007.
- Shuping Shi and Jun Yu. Volatility puzzle: Long memory or antipersistency. *Management Science*, 69(7):3861–3883, 2023.
- Fabian Sobotka and Thomas Kneib. Geoadditive expectile regression. *Computational Statistics & Data Analysis*, 56(4):755–767, 2012.
- Fabian Sobotka, Göran Kauermann, Linda Schulze Waltrup, and Thomas Kneib. On confidence intervals for semiparametric expectile regression. *Statistics and Computing*, 23: 135–148, 2013.
- So Young Sohn and Seung Hwan Lee. Sensitivity analysis for output performance measures in long-range dependent queueing system. *Computers & Operations Research*, 31(9):1527–1536, 2004.
- Paul Speckman. Kernel Smoothing in Partial Linear Models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 50(3):413–436, 1988.
- Natalie M. Steiger and James R. Wilson. An improved batch means procedure for simulation output analysis. *Management Science*, 48(12):1569–1586, 2002. ISSN 00251909, 15265501. URL <http://www.jstor.org/stable/822524>.
- Xiaoqian Sun, Jinhong You, Gemai Chen, and Xian Zhou. Convergence rates of estimators in partial linear regression models with $ma(\infty)$ error process. *Communications in Statistics - Theory and Methods*, 31(12):2251–2273, 2002.

- James W. Taylor. Estimating Value at Risk and Expected Shortfall Using Expectiles. *Journal of Financial Econometrics*, 6(2):231–252, 02 2008. ISSN 1479-8409. doi:10.1093/jjfinec/nbn001. URL <https://doi.org/10.1093/jjfinec/nbn001>.
- John B. Taylor. Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy*, 39:195–214, 1993.
- Howell Tong. *Non-Linear Time Series: A Dynamical System Approach*. Oxford University Press, 1990.
- Ngoc M Tran, Petra Burdejová, Maria Ospienko, and Wolfgang K Härdle. Principal component analysis in an asymmetric norm. *Journal of Multivariate Analysis*, 171:1–21, 2019.
- Joel A Tropp et al. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015.
- Stefanos Tsimenidis, Thomas Lagkas, and Konstantinos Rantos. Deep learning in iot intrusion detection. *Journal of network and systems management*, 30(8):1–40, 2022.
- Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.
- Linda Schulze Waltrup, Fabian Sobotka, Thomas Kneib, and Göran Kauermann. Expectile and quantile regression—david and goliath? *Statistical Modelling*, 15(5):433–456, 2015.
- Ruodu Wang and Ričardas Zitikis. An axiomatic foundation for the expected shortfall. *Management Science*, 67(3):1413–1429, 2021.
- Chunzan Wu. More unequal income but less progressive taxation. *Journal of Monetary Economics*, 117:949–968, 2021.
- Liming Wu. Large and moderate deviations and exponential convergence for stochastic damping hamiltonian systems. *Stochastic processes and their applications*, 91(2):205–238, 2001.
- Wei Biao Wu. On the Bahadur representation of sample quantiles for dependent sequences. *The Annals of Statistics*, 33(4):1934–1963, 2005a.
- Wei Biao Wu. Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences*, 102(40):14150–14154, 2005b.
- Wei Biao Wu. Unit root testing for functionals of linear processes. *Econometric Theory*, 22(1):1–14, 2006.
- Wei Biao Wu. Strong invariance principles for dependent random variables. *The Annals of Probability*, 35(6):2294–2320, 2007a. ISSN 0091-1798. doi:10.1214/009117907000000060. URL <https://doi.org/10.1214/009117907000000060>.
- Wei Biao Wu. M-estimation of linear models with dependent errors. *The Annals of Statistics*, 35(2):495–521, 2007b.

- Wei Biao Wu. Asymptotic theory for stationary processes. *Statistics and Its Interface*, 4(2): 207–226, 2011.
- Wei Biao Wu and Xiaofeng Shao. Limit theorems for iterated random functions. *Journal of Applied Probability*, 41(2):425–436, 2004.
- Wei Biao Wu and Michael Woodroffe. A central limit theorem for iterated random functions. *Journal of applied probability*, 37(3):748–755, 2000.
- Wei Biao Wu and Zhibiao Zhao. Inference of Trends in Time Series. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 69(3):391–410, 05 2007a. ISSN 1369-7412. doi:10.1111/j.1467-9868.2007.00594.x. URL <https://doi.org/10.1111/j.1467-9868.2007.00594.x>.
- Wei Biao Wu and Zhibiao Zhao. Inference of trends in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(3):391–410, 2007b.
- Wei Biao Wu and Zhibiao Zhao. Inference of trends in time series. *Journal of the Royal Statistical Society: Series B*, 69:391–410, 2007c.
- Wei Biao Wu and Zibiao Zhao. Inference of trends in time series. *Journal of the Royal Statistical Society: Series B*, 69:391–410, 2007d.
- Zhaowei Xi, Yu Zhou, Dai Zhang, Kai Gao, Chen Sun, Jiamin Cao, Yangyang Wang, Mingwei Xu, and Jianping Wu. Newton: Intent-driven network traffic monitoring. *IEEE/ACM Transactions on Networking*, 30(2):939–952, 2022. doi:10.1109/TNET.2021.3128557.
- Huiliang Xie and Jian Huang. SCAD-penalized regression in high-dimensional partially linear models. *The Annals of Statistics*, 37(2):673–696, April 2009.
- Lei Yao. *Queuing analysis and control of long range dependent traffic: applications to internet traffic engineering*. The George Washington University, 2003.
- Qiwli Yao and Howell Tong. Asymmetric least squares regression estimation: a nonparametric approach. *Journal of nonparametric statistics*, 6(2-3):273–292, 1996.
- Yan Yu and David Ruppert. Penalized Spline Estimation for Partially Linear Single-Index Models. *Journal of the American Statistical Association*, 97(460):1042–1054, 2002.
- Naiming Yuan, Christian LE Franzke, Feilin Xiong, Zuntao Fu, and Wenjie Dong. The impact of long-term memory on the climate response to greenhouse gas emissions. *npj Climate and Atmospheric Science*, 5(1):70, 2022.
- Tsvetelin S Zaeviski and Dragomir C Nedeltchev. From basel iii to basel iv and beyond: Expected shortfall and expectile risk measures. *International Review of Financial Analysis*, 87:102645, 2023.

- Ting Zhang and Wei Biao Wu. Time-varying nonlinear regression models: Nonparametric estimation and model selection. *The Annals of Statistics*, 43(2):741 – 768, 2015.
- Ting Zhang, Hwai-Chung Ho, Martin Wendler, and Wei Biao Wu. Block sampling under strong dependence. *Stochastic Processes and their Applications*, 123(6):2323–2339, 2013.
- Zhibiao Zhao and Wei Biao Wu. Kernel Quantile Regression for Nonlinear Stochastic Models. *Technical Report*, (No. 572):29, 2006.
- Zhibiao Zhao and Wei Biao Wu. Confidence Bands in Nonparametric Time Series Regression. *The Annals of Statistics*, 36(4):1854 – 1878, 2008a.
- Zhibiao Zhao and Wei Biao Wu. Confidence bands in nonparametric time series regression. *The Annals of Statistics*, 36(4):1854 – 1878, 2008b.
- Zhibiao Zhao and Wei Biao Wu. Confidence bands in nonparametric time series regression. *The Annals of Statistics*, 36:1854–1878, June 2008c.
- Zhou. Zhou and Wei Biao Wu. Simultaneous inference of linear models with time varying coefficients. *Journal of the Royal Statistical Society: Series B*, 72:513–531, 2010a.
- Zhou Zhou and Wei Biao Wu. Simultaneous inference of linear models with time varying coefficients. *Journal of Royal Statistical Society: Series B*, 72:513–531, 2010b.
- Ying Zhu, Zhuqing Yu, and Guang Cheng. High Dimensional Inference in Partially Linear Models. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, pages 2760–2769. PMLR, 2019.
- Johanna F Ziegel. Coherence and elicibility. *Mathematical Finance*, 26(4):901–918, 2016.
- Tobias Zwingmann and Hajo Holzmann. Weak convergence of quantile and expectile processes under general assumptions. *Bernoulli*, 26(1):323 – 351, 2020. doi:10.3150/19-BEJ1127. URL <https://doi.org/10.3150/19-BEJ1127>.

APPENDIX A

APPENDIX: SUPPLEMENTARY MATERIALS FOR CHAPTER

2

A.1 Additional Simulation

A.1.1 Simulations on Queueing processes

In this appendix section, we will present some simulation results of 95%-CIs of M/M/1, M/H₂/1 and M/G/1 processes.

Table A.1: Empirical coverage probabilities of two-sided 95% CIs of expectiles for the M/M/1 process using the d -fold method with sample size n and expectile level α

d	n	$\alpha = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	100	0.950	0.953	0.952	0.952	0.951	0.951	0.946	0.948	0.950
	200	0.950	0.950	0.950	0.953	0.955	0.953	0.953	0.952	0.947
	500	0.951	0.949	0.952	0.954	0.953	0.946	0.950	0.950	0.949
	1000	0.951	0.953	0.951	0.951	0.951	0.950	0.946	0.949	0.948
	5000	0.952	0.948	0.952	0.950	0.952	0.950	0.946	0.951	0.949
3	100	0.947	0.948	0.948	0.949	0.948	0.947	0.945	0.944	0.942
	200	0.951	0.950	0.948	0.950	0.949	0.949	0.948	0.948	0.945
	500	0.950	0.953	0.955	0.953	0.954	0.954	0.953	0.952	0.950
	1000	0.953	0.951	0.952	0.954	0.951	0.953	0.950	0.950	0.951
	5000	0.951	0.951	0.951	0.953	0.952	0.951	0.952	0.951	0.952
4	100	0.950	0.950	0.949	0.950	0.949	0.948	0.946	0.942	0.938
	200	0.951	0.951	0.952	0.952	0.952	0.950	0.948	0.948	0.945
	500	0.951	0.951	0.950	0.951	0.949	0.948	0.946	0.948	0.947
	1000	0.952	0.950	0.951	0.951	0.950	0.949	0.947	0.946	0.946
	5000	0.950	0.950	0.952	0.950	0.949	0.950	0.949	0.951	0.951
5	100	0.951	0.953	0.952	0.949	0.947	0.946	0.945	0.940	0.932
	200	0.947	0.948	0.948	0.948	0.947	0.946	0.945	0.943	0.939
	500	0.949	0.949	0.948	0.949	0.949	0.949	0.948	0.947	0.944
	1000	0.950	0.949	0.946	0.947	0.947	0.948	0.947	0.947	0.948
	5000	0.951	0.953	0.951	0.952	0.953	0.951	0.950	0.950	0.950
6	100	0.948	0.950	0.949	0.949	0.947	0.946	0.944	0.938	0.927
	200	0.951	0.951	0.949	0.949	0.948	0.946	0.944	0.942	0.935
	500	0.948	0.950	0.948	0.948	0.949	0.949	0.949	0.948	0.945
	1000	0.950	0.949	0.951	0.949	0.948	0.947	0.947	0.946	0.947
	5000	0.950	0.951	0.952	0.951	0.951	0.950	0.949	0.949	0.949
10	200	0.949	0.950	0.950	0.949	0.949	0.947	0.942	0.935	0.923
	1000	0.949	0.950	0.946	0.945	0.945	0.944	0.943	0.942	0.940
	5000	0.952	0.951	0.952	0.951	0.950	0.951	0.950	0.949	0.950
	10000	0.949	0.950	0.949	0.947	0.946	0.947	0.948	0.948	0.949
20	1000	0.945	0.948	0.950	0.948	0.947	0.947	0.945	0.941	0.935
	5000	0.950	0.950	0.951	0.953	0.954	0.954	0.952	0.952	0.950
	10000	0.950	0.949	0.948	0.949	0.950	0.950	0.950	0.950	0.949
	100000	0.946	0.948	0.948	0.947	0.948	0.946	0.948	0.948	0.947
100	5000	0.900	0.935	0.945	0.950	0.950	0.948	0.944	0.936	0.910
	10000	0.925	0.943	0.947	0.949	0.950	0.948	0.946	0.942	0.933
	100000	0.944	0.945	0.946	0.948	0.946	0.946	0.945	0.945	0.946

Table A.2: Halfwidths (multiplied by $\sqrt{n}$) of two-sided 95% CIs of expectiles for the M/M/1 process using the d -fold method with sample size n and expectile level α

d	n	$\alpha = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	100	46.337	61.643	74.600	87.509	101.63	118.26	139.81	172.12	237.20
	200	45.901	61.271	74.251	87.181	101.19	117.63	138.93	170.97	235.55
	500	45.520	60.870	73.866	86.683	100.63	117.11	138.77	171.59	237.99
	1000	45.863	61.275	74.272	87.143	101.21	117.84	139.40	171.96	237.48
	5000	45.756	61.092	74.254	87.326	101.55	118.25	139.97	172.67	238.08
3	100	17.537	23.274	28.133	32.944	38.164	44.324	52.343	64.371	88.303
	200	17.194	22.864	27.665	32.434	37.624	43.764	51.759	63.753	87.871
	500	17.196	22.993	27.920	32.796	38.107	44.390	52.565	64.867	89.496
	1000	17.247	23.031	27.933	32.799	38.085	44.328	52.449	64.616	89.202
	5000	17.244	23.086	28.070	33.011	38.387	44.734	52.982	65.367	90.191
4	100	13.496	17.833	21.526	25.196	29.192	33.901	40.002	49.124	67.284
	200	13.328	17.733	21.458	25.152	29.162	33.886	40.020	49.225	67.721
	500	13.193	17.605	21.344	25.055	29.096	33.877	40.108	49.491	68.404
	1000	13.240	17.655	21.404	25.134	29.198	34.005	40.258	49.649	68.629
	5000	13.104	17.540	21.313	25.069	29.163	34.006	40.309	49.775	68.797
5	100	12.082	15.896	19.148	22.383	25.907	30.067	35.470	43.534	59.513
	200	11.822	15.691	18.988	22.259	25.811	30.000	35.440	43.594	59.967
	500	11.852	15.804	19.142	22.445	26.036	30.284	35.821	44.153	60.917
	1000	11.759	15.706	19.056	22.391	26.023	30.317	35.911	44.315	61.285
	5000	11.688	15.669	19.052	22.420	26.100	30.454	36.119	44.624	61.743
6	100	11.340	14.864	17.880	20.878	24.142	27.992	32.982	40.415	55.017
	200	11.267	14.900	17.987	21.050	24.385	28.325	33.451	41.140	56.523
	500	11.125	14.839	17.991	21.114	24.511	28.527	33.756	41.617	57.388
	1000	10.995	14.673	17.798	20.906	24.290	28.294	33.507	41.349	57.173
	5000	10.936	14.630	17.770	20.896	24.308	28.351	33.623	41.547	57.485
10	200	10.183	13.375	16.104	18.822	21.787	25.291	29.839	36.646	50.163
	1000	9.952	13.263	16.075	18.868	21.911	25.511	30.203	37.259	51.492
	5000	9.834	13.174	16.012	18.828	21.894	25.519	30.241	37.341	51.638
	10000	9.860	13.180	15.988	18.777	21.820	25.423	30.122	37.197	51.483
20	1000	9.444	12.525	15.147	17.751	20.590	23.950	28.325	34.901	48.108
	5000	9.293	12.419	15.065	17.687	20.543	23.921	28.318	34.928	48.234
	10000	9.288	12.418	15.066	17.692	20.552	23.937	28.348	34.979	48.357
	100000	9.272	12.402	15.048	17.672	20.529	23.908	28.313	34.951	48.366
100	5000	9.020	11.969	14.476	16.967	19.684	22.900	27.089	33.388	46.067
	10000	8.959	11.932	14.457	16.965	19.700	22.936	27.154	33.500	46.286
	100000	8.886	11.882	14.418	16.935	19.675	22.918	27.145	33.509	46.358

Table A.3: Empirical coverage probabilities of two-sided 95% CIs of expectiles for the $M/H_2/1$ process using the d -fold method with sample size n and expectile level α

d	n	$\alpha = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	100	0.951	0.950	0.951	0.948	0.951	0.953	0.954	0.952	0.952
	200	0.948	0.949	0.951	0.950	0.954	0.950	0.950	0.953	0.952
	500	0.952	0.948	0.951	0.950	0.949	0.951	0.953	0.950	0.951
	1000	0.949	0.948	0.949	0.947	0.949	0.952	0.950	0.950	0.951
	5000	0.952	0.952	0.951	0.954	0.953	0.952	0.954	0.951	0.951
3	100	0.949	0.948	0.949	0.949	0.947	0.948	0.949	0.948	0.945
	200	0.948	0.950	0.948	0.946	0.947	0.949	0.951	0.951	0.947
	500	0.947	0.948	0.946	0.947	0.948	0.947	0.947	0.950	0.948
	1000	0.946	0.947	0.950	0.952	0.952	0.951	0.950	0.949	0.949
	5000	0.951	0.949	0.950	0.951	0.952	0.954	0.954	0.952	0.952
4	100	0.952	0.951	0.947	0.949	0.949	0.947	0.946	0.944	0.940
	200	0.955	0.955	0.954	0.957	0.955	0.953	0.949	0.948	0.946
	500	0.951	0.952	0.950	0.950	0.950	0.950	0.950	0.949	0.947
	1000	0.947	0.947	0.947	0.945	0.944	0.946	0.946	0.948	0.944
	5000	0.950	0.951	0.951	0.951	0.950	0.948	0.948	0.949	0.952
5	100	0.954	0.952	0.955	0.952	0.952	0.949	0.947	0.943	0.937
	200	0.952	0.952	0.951	0.951	0.951	0.952	0.952	0.948	0.945
	500	0.950	0.950	0.950	0.951	0.948	0.947	0.947	0.948	0.947
	1000	0.948	0.948	0.948	0.947	0.947	0.947	0.946	0.944	0.944
	5000	0.951	0.949	0.948	0.949	0.949	0.949	0.950	0.950	0.949
6	100	0.948	0.948	0.949	0.946	0.946	0.944	0.941	0.937	0.929
	200	0.947	0.950	0.951	0.951	0.950	0.949	0.948	0.943	0.937
	500	0.949	0.948	0.948	0.947	0.947	0.949	0.947	0.946	0.944
	1000	0.951	0.949	0.948	0.949	0.947	0.945	0.945	0.945	0.946
	5000	0.951	0.950	0.949	0.948	0.948	0.950	0.950	0.951	0.950
10	200	0.951	0.953	0.952	0.951	0.950	0.948	0.945	0.941	0.928
	1000	0.948	0.949	0.947	0.946	0.944	0.943	0.942	0.943	0.940
	5000	0.949	0.949	0.951	0.950	0.949	0.950	0.951	0.949	0.949
	10000	0.950	0.949	0.950	0.949	0.948	0.947	0.948	0.950	0.949
20	1000	0.942	0.947	0.948	0.947	0.947	0.946	0.945	0.940	0.934
	5000	0.947	0.949	0.950	0.949	0.949	0.948	0.949	0.950	0.948
	10000	0.949	0.951	0.951	0.951	0.952	0.952	0.952	0.952	0.948
	100000	0.950	0.951	0.950	0.952	0.950	0.951	0.950	0.952	0.949
100	5000	0.890	0.930	0.944	0.949	0.950	0.949	0.942	0.931	0.909
	10000	0.919	0.941	0.947	0.948	0.950	0.950	0.948	0.943	0.930
	100000	0.943	0.948	0.950	0.950	0.949	0.949	0.950	0.949	0.946

Table A.4: Halfwidths (multiplied by $\sqrt{n}$) of two-sided 95% CIs of expectiles for the M/H₂/1 process using the d -fold method with sample size n and expectile level α

d	n	$\alpha = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	100	65.631	94.035	117.06	139.13	162.72	190.30	225.89	278.73	384.04
	200	64.267	92.269	115.08	137.15	160.87	188.60	224.59	278.70	387.64
	500	63.897	92.076	115.00	137.18	161.05	189.13	225.40	279.54	387.99
	1000	63.404	91.147	113.83	135.78	159.40	187.17	223.01	276.31	383.20
	5000	64.424	93.121	116.35	138.78	162.67	190.63	226.82	281.04	390.96
3	100	24.608	34.973	43.419	51.600	60.374	70.655	83.970	103.88	143.34
	200	24.192	34.666	43.195	51.452	60.331	70.725	84.175	104.25	144.17
	500	24.068	34.673	43.328	51.714	60.698	71.183	84.673	104.78	144.98
	1000	23.964	34.533	43.157	51.475	60.376	70.793	84.257	104.34	144.47
	5000	23.869	34.482	43.140	51.506	60.468	70.954	84.512	104.81	145.43
4	100	19.122	27.092	33.595	39.883	46.603	54.434	64.533	79.565	109.26
	200	18.847	26.954	33.559	39.955	46.792	54.768	65.058	80.441	111.20
	500	18.588	26.750	33.360	39.735	46.556	54.536	64.845	80.244	111.12
	1000	18.449	26.576	33.197	39.591	46.421	54.387	64.667	80.061	110.98
	5000	18.514	26.766	33.465	39.933	46.841	54.913	65.338	80.916	112.15
5	100	17.167	24.197	29.962	35.548	41.524	48.492	57.461	70.783	96.916
	200	16.740	23.863	29.673	35.300	41.330	48.381	57.488	71.100	98.309
	500	16.607	23.853	29.736	35.408	41.470	48.554	57.695	71.356	98.730
	1000	16.477	23.725	29.613	35.287	41.352	48.436	57.593	71.315	98.942
	5000	16.431	23.712	29.628	35.343	41.461	48.619	57.876	71.725	99.523
6	100	16.036	22.437	27.706	32.819	38.293	44.688	52.933	65.167	89.119
	200	15.847	22.538	27.996	33.283	38.949	45.571	54.111	66.838	92.142
	500	15.607	22.406	27.930	33.278	39.005	45.698	54.329	67.199	92.950
	1000	15.401	22.142	27.629	32.933	38.617	45.268	53.875	66.757	92.541
	5000	15.411	22.254	27.833	33.229	39.006	45.765	54.499	67.545	93.542
10	200	14.429	20.301	25.129	29.826	34.867	40.759	48.356	59.671	82.006
	1000	13.955	20.030	24.977	29.748	34.849	40.807	48.506	60.020	83.152
	5000	13.864	19.997	24.993	29.824	35.000	41.055	48.881	60.583	84.029
	10000	13.867	19.999	24.977	29.788	34.942	40.972	48.759	60.392	83.719
20	1000	13.247	18.905	23.527	27.999	32.785	38.379	45.604	56.401	78.029
	5000	13.056	18.806	23.489	28.018	32.867	38.535	45.858	56.812	78.784
	10000	13.028	18.770	23.434	27.946	32.783	38.438	45.740	56.644	78.499
	100000	13.039	18.822	23.526	28.065	32.920	38.590	45.913	56.863	78.830
100	5000	12.732	18.156	22.594	26.892	31.500	36.889	43.852	54.252	75.058
	10000	12.600	18.079	22.549	26.879	31.520	36.947	43.958	54.431	75.412
	100000	12.477	17.999	22.493	26.835	31.485	36.919	43.938	54.434	75.514

Table A.5: Empirical coverage probabilities of two-sided 95% CIs of expectiles for the M/G/1 process with lognormal service time using the d -fold method with sample size n and expectile level α

d	n	$\alpha = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	100	0.948	0.944	0.941	0.942	0.941	0.939	0.936	0.931	0.926
	200	0.948	0.947	0.944	0.942	0.940	0.941	0.940	0.938	0.930
	500	0.947	0.945	0.944	0.942	0.944	0.947	0.946	0.941	0.940
	1000	0.953	0.951	0.950	0.948	0.948	0.949	0.949	0.947	0.944
	5000	0.950	0.949	0.948	0.951	0.950	0.949	0.949	0.949	0.951
3	100	0.926	0.923	0.919	0.919	0.916	0.908	0.902	0.892	0.869
	200	0.933	0.931	0.929	0.927	0.923	0.919	0.917	0.911	0.899
	500	0.945	0.939	0.937	0.937	0.935	0.932	0.928	0.926	0.916
	1000	0.948	0.948	0.948	0.948	0.948	0.947	0.944	0.940	0.936
	5000	0.948	0.948	0.948	0.947	0.946	0.946	0.946	0.946	0.945
4	100	0.920	0.912	0.903	0.896	0.889	0.880	0.872	0.860	0.829
	200	0.929	0.922	0.918	0.914	0.909	0.905	0.898	0.889	0.868
	500	0.938	0.935	0.932	0.929	0.926	0.924	0.920	0.914	0.906
	1000	0.944	0.942	0.942	0.940	0.937	0.935	0.932	0.929	0.920
	5000	0.946	0.943	0.942	0.941	0.940	0.940	0.941	0.941	0.939
5	100	0.915	0.904	0.897	0.888	0.880	0.870	0.857	0.839	0.798
	200	0.921	0.913	0.908	0.904	0.900	0.894	0.886	0.874	0.848
	500	0.935	0.933	0.930	0.926	0.921	0.918	0.914	0.909	0.896
	1000	0.938	0.936	0.934	0.933	0.930	0.926	0.924	0.920	0.913
	5000	0.945	0.944	0.946	0.945	0.945	0.945	0.945	0.944	0.942
6	100	0.912	0.899	0.890	0.882	0.874	0.863	0.852	0.829	0.775
	200	0.921	0.913	0.907	0.902	0.897	0.890	0.881	0.867	0.837
	500	0.933	0.930	0.925	0.922	0.917	0.914	0.908	0.901	0.886
	1000	0.938	0.936	0.934	0.932	0.930	0.927	0.923	0.917	0.906
	5000	0.946	0.945	0.944	0.944	0.944	0.945	0.943	0.943	0.939
10	200	0.917	0.904	0.896	0.889	0.879	0.871	0.860	0.841	0.797
	1000	0.933	0.928	0.926	0.923	0.921	0.919	0.914	0.908	0.897
	5000	0.945	0.944	0.943	0.943	0.942	0.943	0.942	0.938	0.934
	10000	0.948	0.947	0.947	0.946	0.946	0.945	0.944	0.943	0.941
20	1000	0.933	0.927	0.924	0.921	0.918	0.914	0.910	0.901	0.880
	5000	0.943	0.940	0.938	0.937	0.936	0.937	0.936	0.935	0.930
	10000	0.945	0.945	0.944	0.943	0.943	0.943	0.943	0.942	0.939
	100000	0.952	0.953	0.954	0.953	0.953	0.953	0.953	0.953	0.952
100	5000	0.941	0.939	0.938	0.937	0.937	0.937	0.935	0.933	0.920
	10000	0.943	0.941	0.942	0.940	0.941	0.942	0.943	0.943	0.942
	100000	0.950	0.951	0.951	0.952	0.951	0.951	0.951	0.951	0.952

Table A.6: Halfwidths (multiplied by $\sqrt{n}$) of two-sided 95% CIs of expectiles for the M/G/1 process with lognormal service time using the d -fold method with sample size n and expectile level α

d	n	$\alpha = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	100	25.188	43.276	63.007	86.456	116.20	156.78	217.78	324.96	585.62
	200	26.182	45.215	66.026	90.782	122.24	165.39	230.78	347.36	640.66
	500	27.390	47.452	69.476	95.712	129.18	175.13	245.05	370.75	694.87
	1000	28.577	49.476	72.363	99.618	134.31	181.96	254.49	385.07	723.21
	5000	29.534	51.309	75.190	103.64	139.88	189.71	265.71	403.10	761.79
3	100	9.205	15.821	23.053	31.635	42.502	57.264	79.317	117.65	208.02
	200	9.555	16.482	24.063	33.095	44.602	60.373	84.236	126.64	232.26
	500	10.191	17.631	25.794	35.519	47.906	64.930	90.839	137.40	257.00
	1000	10.530	18.256	26.735	36.838	49.708	67.394	94.339	142.89	268.70
	5000	10.997	19.080	27.950	38.535	52.050	70.660	99.064	150.45	284.43
4	100	6.989	11.994	17.460	23.947	32.151	43.270	59.798	88.220	153.08
	200	7.343	12.662	18.482	25.407	34.219	46.278	64.476	96.651	175.50
	500	7.760	13.453	19.700	27.153	36.657	49.724	69.619	105.38	196.94
	1000	7.998	13.865	20.302	27.976	37.761	51.221	71.743	108.75	204.69
	5000	8.393	14.578	21.367	29.471	39.819	54.068	75.823	115.16	217.82
5	100	6.177	10.603	15.442	21.182	28.428	38.211	52.661	77.192	131.34
	200	6.455	11.150	16.299	22.433	30.238	40.912	56.992	85.296	153.67
	500	6.850	11.870	17.386	23.970	32.378	43.950	61.578	93.257	174.08
	1000	7.090	12.291	18.005	24.828	33.539	45.535	63.836	96.864	182.42
	5000	7.458	12.947	18.975	26.179	35.384	48.064	67.435	102.50	194.01
6	100	5.724	9.828	14.316	19.637	26.338	35.361	48.619	70.892	118.64
	200	6.055	10.444	15.263	21.007	28.313	38.290	53.284	79.537	141.91
	500	6.426	11.144	16.334	22.537	30.463	41.377	58.004	87.851	163.61
	1000	6.595	11.441	16.765	23.123	31.245	42.435	59.517	90.378	170.42
	5000	6.983	12.124	17.771	24.521	33.149	45.043	63.222	96.169	182.31
10	200	5.360	9.259	13.552	18.676	25.182	34.022	47.167	69.629	119.50
	1000	5.876	10.207	14.981	20.700	28.027	38.146	53.623	81.608	153.72
	5000	6.231	10.830	15.890	21.943	29.690	40.385	56.766	86.529	164.69
	10000	6.324	10.994	16.129	22.271	30.128	40.969	57.561	87.686	166.77
20	1000	5.439	9.476	13.949	19.328	26.242	35.809	50.448	76.783	142.83
	5000	5.784	10.074	14.807	20.485	27.770	37.857	53.362	81.664	156.39
	10000	5.897	10.264	15.076	20.844	28.237	38.463	54.165	82.794	158.47
	100000	6.063	10.548	15.480	21.377	28.916	39.310	55.208	84.043	159.61
100	5000	5.359	9.407	13.935	19.429	26.549	36.488	51.845	79.795	150.96
	10000	5.494	9.628	14.233	19.806	27.017	37.082	52.682	81.372	157.17
	100000	5.763	10.044	14.764	20.423	27.681	37.725	53.159	81.346	156.19

A.1.2 Simulation on High Confidence Level Inference

This appendix presents the simulation results of two-sided 99% expectile CIs for AR(1) and TAR(1) processes. Our findings exhibit similarity to those previously attained.

A.2 Technical Proofs

A.2.1 Proof of Theorem 2.2.1

Proof of Theorem 2.2.1. Note that the expectile estimator $\hat{\theta}_\alpha$ in (2.2) is a nonlinear function of $X_1, \dots, X_n$. Thus it is not convenient to analyze it directly. Here we shall establish a Bahadur representation which gives an asymptotic linear approximation. Based on the latter result the central limit theorem follows. To this end, we shall apply Theorem 1 in Wu [2007b]. First consider the case with $d = 1$. For the sake of completeness, we will verify the conditions (A1)-(A4) in Theorem 1 in Wu [2007b] individually.

(A1) The asymmetric quadratic loss function ρ_α is convex. By the definition of α th expectile, $\mathbb{E}[g_\alpha(X_1 - \theta_\alpha)] = 0$. We also have $\|g_\alpha(X_1 - \theta_\alpha)\|^2 > 0$ unless $\mathbb{P}(X_1 = \theta_\alpha) = 1$, which contradicts with the condition that F is continuous at θ_α .

(A2) Let $\varphi(u) = \mathbb{E}g_\alpha(X_1 - \theta_\alpha + u)$. Then by integration by parts,

$$\varphi(u) = \alpha \int_{\theta_\alpha - u}^{\infty} 1 - F(x) dx + (1 - \alpha) \int_{-\infty}^{\theta_\alpha - u} F(x) dx.$$

Since F is continuous at θ_α , we have

$$\varphi'(0) = \alpha(1 - F(\theta_\alpha)) + (1 - \alpha)F(\theta_\alpha) = \tau_\alpha > 0. \quad (\text{A.1})$$

(A3) Clearly $g_\alpha(\cdot)$ is strictly increasing and the function $\mu(u) = \|g_\alpha(X_1 + u - \theta_\alpha) - g_\alpha(X_1 - \theta_\alpha)\| \leq |u|$ is continuous at $u = 0$.

Table A.7: Empirical coverage probabilities of two-sided 99% expectile CIs for the AR(1) process with $t(3)$ innovations using the d -fold method with sample size n and expectile level α with 10,000 replications

d	n	$\alpha = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	100	0.989	0.991	0.989	0.990	0.992	0.990	0.991	0.989	0.990
	200	0.989	0.989	0.991	0.991	0.991	0.991	0.989	0.990	0.989
	500	0.990	0.991	0.992	0.990	0.990	0.990	0.990	0.990	0.990
	1000	0.990	0.990	0.990	0.991	0.991	0.991	0.991	0.989	0.990
	5000	0.991	0.990	0.991	0.989	0.990	0.990	0.990	0.991	0.991
3	100	0.985	0.990	0.989	0.989	0.989	0.990	0.990	0.989	0.987
	200	0.990	0.990	0.990	0.991	0.990	0.991	0.990	0.990	0.989
	500	0.990	0.991	0.991	0.991	0.992	0.992	0.990	0.988	0.988
	1000	0.988	0.990	0.990	0.991	0.990	0.990	0.990	0.991	0.990
	5000	0.989	0.991	0.990	0.991	0.991	0.991	0.990	0.989	0.990
4	100	0.979	0.985	0.987	0.989	0.990	0.989	0.989	0.988	0.983
	200	0.985	0.990	0.990	0.991	0.991	0.991	0.988	0.989	0.984
	500	0.988	0.990	0.990	0.989	0.989	0.988	0.987	0.988	0.986
	1000	0.988	0.989	0.989	0.989	0.991	0.990	0.990	0.991	0.987
	5000	0.989	0.988	0.990	0.991	0.993	0.993	0.991	0.990	0.988
5	100	0.974	0.983	0.988	0.989	0.989	0.989	0.988	0.985	0.974
	200	0.980	0.988	0.990	0.992	0.991	0.990	0.990	0.988	0.981
	500	0.986	0.992	0.991	0.990	0.990	0.989	0.989	0.988	0.984
	1000	0.985	0.989	0.990	0.990	0.990	0.990	0.990	0.988	0.986
	5000	0.990	0.990	0.990	0.991	0.990	0.990	0.990	0.990	0.989
6	100	0.965	0.982	0.987	0.989	0.988	0.988	0.987	0.981	0.965
	200	0.976	0.986	0.988	0.988	0.990	0.989	0.989	0.986	0.975
	500	0.982	0.990	0.991	0.991	0.991	0.990	0.988	0.986	0.981
	1000	0.985	0.989	0.990	0.990	0.990	0.989	0.988	0.987	0.984
	5000	0.990	0.990	0.991	0.992	0.992	0.992	0.990	0.989	0.987
10	200	0.952	0.979	0.986	0.989	0.990	0.989	0.986	0.979	0.956
	1000	0.980	0.988	0.989	0.989	0.989	0.988	0.987	0.983	0.978
	5000	0.987	0.990	0.990	0.990	0.990	0.990	0.989	0.989	0.986
	10000	0.988	0.989	0.990	0.991	0.991	0.990	0.990	0.990	0.988
20	1000	0.966	0.980	0.986	0.988	0.988	0.988	0.985	0.981	0.969
	5000	0.986	0.989	0.991	0.991	0.991	0.990	0.990	0.988	0.984
	10000	0.986	0.989	0.990	0.991	0.990	0.991	0.990	0.990	0.988
	100000	0.989	0.990	0.990	0.990	0.990	0.991	0.990	0.991	0.990
100	5000	0.958	0.976	0.985	0.989	0.989	0.988	0.984	0.977	0.956
	10000	0.977	0.984	0.986	0.989	0.990	0.989	0.989	0.985	0.979
	100000	0.990	0.991	0.990	0.990	0.990	0.990	0.991	0.991	0.990

Table A.8: Halfwidths (multiplied by $\sqrt{n}$) of two-sided 99% expectile CIs for the AR(1) process with innovations $t(3)$ using the d -fold method with sample size n and expectile level α with 10,000 replications

d	n	$\alpha = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	100	252.68	199.13	179.55	171.37	169.35	172.45	182.15	203.96	261.26
	200	275.80	209.50	185.62	175.10	171.57	173.88	183.24	205.83	270.05
	500	283.08	212.21	187.43	176.97	174.09	177.26	187.66	211.95	282.00
	1000	286.38	213.38	188.74	178.45	175.55	178.52	189.13	214.04	286.21
	5000	294.06	218.39	192.29	181.08	177.65	180.48	191.12	216.52	290.85
3	100	42.378	33.971	30.707	29.280	28.906	29.403	30.960	34.427	43.272
	200	46.234	35.561	31.652	29.965	29.472	29.956	31.628	35.544	46.309
	500	48.411	36.222	31.993	30.215	29.708	30.212	31.957	36.085	48.014
	1000	49.650	36.888	32.503	30.641	30.085	30.591	32.397	36.657	49.171
	5000	50.139	37.259	32.862	30.988	30.431	30.939	32.777	37.168	50.065
4	100	25.083	20.389	18.526	17.707	17.490	17.776	18.680	20.664	25.584
	200	27.788	21.544	19.227	18.221	17.929	18.220	19.228	21.553	27.791
	500	29.625	22.203	19.613	18.508	18.182	18.480	19.550	22.077	29.335
	1000	30.152	22.436	19.822	18.731	18.425	18.746	19.848	22.455	30.122
	5000	30.725	22.801	20.112	18.965	18.620	18.922	20.034	22.699	30.580
5	100	19.597	16.146	14.732	14.099	13.929	14.150	14.844	16.342	19.958
	200	22.095	17.298	15.466	14.659	14.422	14.654	15.457	17.291	22.114
	500	23.749	17.836	15.751	14.859	14.595	14.833	15.691	17.722	23.510
	1000	24.218	18.005	15.904	15.028	14.787	15.054	15.954	18.073	24.276
	5000	24.678	18.272	16.099	15.177	14.905	15.155	16.057	18.213	24.586
6	100	16.757	13.942	12.756	12.222	12.081	12.270	12.857	14.113	17.062
	200	19.313	15.209	13.617	12.912	12.705	12.907	13.606	15.196	19.306
	500	21.033	15.839	13.993	13.203	12.969	13.181	13.943	15.742	20.827
	1000	21.511	15.959	14.064	13.268	13.039	13.265	14.056	15.940	21.451
	5000	21.961	16.233	14.295	13.475	13.234	13.458	14.259	16.175	21.850
10	200	14.974	12.148	10.986	10.459	10.305	10.463	10.996	12.167	15.006
	1000	17.763	13.182	11.595	10.925	10.731	10.917	11.577	13.148	17.686
	5000	18.184	13.389	11.781	11.105	10.909	11.098	11.765	13.359	18.134
	10000	18.175	13.404	11.798	11.121	10.925	11.116	11.791	13.398	18.170
20	1000	15.545	11.688	10.282	9.680	9.507	9.677	10.275	11.672	15.502
	5000	16.443	11.981	10.502	9.885	9.709	9.882	10.492	11.963	16.403
	10000	16.410	12.003	10.539	9.927	9.752	9.926	10.537	12.002	16.417
	100000	16.358	12.054	10.596	9.981	9.802	9.973	10.580	12.024	16.300
100	5000	15.102	11.108	9.671	9.061	8.885	9.057	9.662	11.089	15.056
	10000	15.650	11.240	9.758	9.146	8.972	9.146	9.759	11.242	15.658
	100000	15.450	11.203	9.808	9.226	9.060	9.224	9.803	11.193	15.431

Table A.9: Empirical coverage probabilities of two-sided 99% CIs of expectiles for the TAR(1) process with innovations $\mathbf{N}(0, 1)$ using the d -fold method with sample size n and expectile level α with 10,000 replications

d	n	$\alpha = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	100	0.991	0.989	0.988	0.991	0.989	0.989	0.990	0.990	0.988
	200	0.991	0.990	0.988	0.989	0.991	0.991	0.991	0.993	0.990
	500	0.989	0.991	0.990	0.989	0.989	0.989	0.990	0.988	0.990
	1000	0.990	0.991	0.991	0.991	0.989	0.990	0.989	0.990	0.991
	5000	0.989	0.989	0.988	0.990	0.990	0.990	0.990	0.990	0.989
3	100	0.989	0.990	0.991	0.990	0.990	0.989	0.990	0.991	0.990
	200	0.990	0.991	0.991	0.990	0.990	0.991	0.991	0.992	0.991
	500	0.991	0.990	0.990	0.991	0.990	0.990	0.988	0.990	0.990
	1000	0.988	0.991	0.991	0.991	0.989	0.989	0.991	0.990	0.989
	5000	0.990	0.988	0.989	0.990	0.991	0.990	0.990	0.990	0.990
4	100	0.988	0.988	0.990	0.989	0.989	0.989	0.990	0.990	0.990
	200	0.990	0.991	0.991	0.990	0.990	0.991	0.992	0.991	0.991
	500	0.990	0.989	0.989	0.990	0.989	0.989	0.989	0.989	0.991
	1000	0.989	0.990	0.990	0.990	0.989	0.989	0.990	0.990	0.990
	5000	0.991	0.992	0.991	0.991	0.991	0.991	0.992	0.990	0.990
5	100	0.985	0.987	0.989	0.989	0.989	0.988	0.990	0.992	0.992
	200	0.986	0.987	0.989	0.990	0.991	0.993	0.991	0.992	0.991
	500	0.989	0.989	0.988	0.989	0.988	0.989	0.989	0.989	0.989
	1000	0.989	0.991	0.991	0.990	0.991	0.991	0.990	0.989	0.989
	5000	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.989
6	100	0.983	0.986	0.988	0.988	0.988	0.990	0.991	0.991	0.991
	200	0.988	0.988	0.989	0.990	0.990	0.990	0.992	0.992	0.991
	500	0.988	0.990	0.991	0.990	0.990	0.989	0.989	0.989	0.989
	1000	0.989	0.991	0.990	0.990	0.990	0.990	0.990	0.990	0.990
	5000	0.991	0.991	0.989	0.989	0.990	0.989	0.990	0.990	0.991
10	200	0.980	0.986	0.987	0.989	0.989	0.990	0.990	0.991	0.990
	1000	0.989	0.990	0.989	0.989	0.991	0.991	0.990	0.990	0.988
	5000	0.991	0.991	0.990	0.989	0.989	0.989	0.989	0.990	0.989
	10000	0.991	0.992	0.992	0.991	0.991	0.990	0.990	0.991	0.990
20	1000	0.983	0.988	0.988	0.989	0.989	0.990	0.989	0.990	0.989
	5000	0.989	0.990	0.990	0.990	0.990	0.990	0.989	0.988	0.988
	10000	0.991	0.990	0.990	0.991	0.991	0.991	0.991	0.990	0.990
	100000	0.990	0.990	0.990	0.991	0.990	0.990	0.991	0.990	0.991
100	5000	0.951	0.975	0.983	0.989	0.989	0.989	0.987	0.984	0.976
	10000	0.973	0.985	0.987	0.990	0.990	0.991	0.989	0.987	0.983
	100000	0.988	0.990	0.990	0.990	0.990	0.989	0.989	0.990	0.990

Table A.10: Halfwidths (multiplied by $\sqrt{n}$) of two-sided 99% CIs of expectiles for the TAR(1) process with innovations $\mathbf{N}(0, 1)$ using the d -fold method with sample size n and expectile level α with 10,000 replications

d	n	$\alpha = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	100	86.279	76.257	71.125	67.749	65.285	63.458	62.330	62.359	65.999
	200	85.522	75.438	70.250	66.798	64.253	62.384	61.146	60.972	64.644
	500	86.689	76.010	70.549	66.920	64.218	62.155	60.760	60.504	64.066
	1000	87.761	77.192	71.740	68.123	65.400	63.296	61.761	61.383	64.811
	5000	86.831	76.222	70.864	67.321	64.658	62.643	61.239	60.948	64.365
3	100	14.717	13.062	12.206	11.638	11.224	10.925	10.746	10.766	11.415
	200	14.912	13.182	12.270	11.659	11.205	10.863	10.630	10.583	11.164
	500	14.882	13.106	12.203	11.605	11.166	10.835	10.608	10.568	11.172
	1000	15.034	13.239	12.315	11.701	11.248	10.901	10.663	10.622	11.234
	5000	15.171	13.346	12.415	11.796	11.333	10.972	10.714	10.641	11.201
4	100	8.852	7.888	7.382	7.045	6.803	6.628	6.525	6.549	6.955
	200	9.023	7.993	7.459	7.105	6.847	6.655	6.530	6.520	6.899
	500	9.137	8.041	7.478	7.107	6.834	6.630	6.494	6.480	6.865
	1000	9.217	8.113	7.543	7.162	6.877	6.658	6.503	6.467	6.826
	5000	9.266	8.154	7.582	7.199	6.912	6.689	6.533	6.497	6.864
5	100	7.047	6.308	5.915	5.654	5.468	5.336	5.266	5.299	5.650
	200	7.219	6.401	5.974	5.690	5.482	5.328	5.227	5.224	5.536
	500	7.348	6.464	6.009	5.709	5.488	5.322	5.210	5.192	5.495
	1000	7.403	6.511	6.052	5.746	5.518	5.343	5.221	5.193	5.483
	5000	7.437	6.539	6.083	5.779	5.551	5.373	5.246	5.212	5.502
6	100	6.119	5.493	5.164	4.947	4.795	4.690	4.637	4.675	4.986
	200	6.397	5.680	5.306	5.059	4.879	4.747	4.663	4.664	4.939
	500	6.497	5.729	5.331	5.068	4.874	4.727	4.628	4.613	4.879
	1000	6.524	5.750	5.351	5.086	4.890	4.740	4.636	4.617	4.879
	5000	6.598	5.803	5.396	5.123	4.920	4.762	4.649	4.618	4.866
10	200	5.135	4.591	4.304	4.116	3.981	3.886	3.833	3.854	4.103
	1000	5.375	4.736	4.407	4.188	4.027	3.904	3.821	3.807	4.029
	5000	5.433	4.781	4.446	4.222	4.055	3.925	3.833	3.811	4.024
	10000	5.420	4.765	4.431	4.209	4.044	3.916	3.825	3.804	4.022
20	1000	4.745	4.199	3.918	3.733	3.597	3.495	3.428	3.422	3.620
	5000	4.836	4.256	3.958	3.760	3.612	3.498	3.418	3.399	3.591
	10000	4.837	4.256	3.958	3.759	3.610	3.496	3.414	3.395	3.589
	100000	4.845	4.259	3.960	3.759	3.610	3.494	3.412	3.393	3.585
100	5000	4.401	3.893	3.633	3.460	3.333	3.238	3.176	3.171	3.357
	10000	4.447	3.921	3.652	3.473	3.341	3.239	3.170	3.158	3.339
	100000	4.488	3.946	3.668	3.484	3.346	3.239	3.163	3.145	3.324

Table A.11: Empirical coverage probabilities of two-sided 99% CIs of expectiles for the TAR(1) process with innovations $t(3)$ using the d -fold method with sample size n and expectile level α with 10,000 replications

d	n	$\alpha = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	100	0.988	0.989	0.988	0.991	0.989	0.992	0.992	0.991	0.990
	200	0.990	0.992	0.992	0.991	0.990	0.990	0.990	0.993	0.990
	500	0.989	0.989	0.989	0.990	0.990	0.990	0.991	0.990	0.991
	1000	0.991	0.990	0.989	0.991	0.992	0.991	0.991	0.991	0.990
	5000	0.989	0.990	0.990	0.991	0.991	0.988	0.989	0.990	0.990
3	100	0.983	0.986	0.985	0.986	0.988	0.989	0.988	0.990	0.989
	200	0.986	0.988	0.988	0.990	0.991	0.991	0.990	0.991	0.990
	500	0.988	0.990	0.991	0.991	0.991	0.992	0.992	0.991	0.989
	1000	0.988	0.989	0.988	0.989	0.989	0.990	0.990	0.990	0.990
	5000	0.989	0.989	0.990	0.990	0.990	0.991	0.992	0.991	0.990
4	100	0.972	0.980	0.980	0.984	0.985	0.988	0.990	0.991	0.989
	200	0.978	0.984	0.985	0.987	0.989	0.989	0.990	0.991	0.988
	500	0.985	0.989	0.988	0.988	0.989	0.989	0.989	0.989	0.987
	1000	0.985	0.987	0.989	0.989	0.989	0.990	0.989	0.991	0.989
	5000	0.988	0.990	0.990	0.989	0.989	0.990	0.991	0.993	0.990
5	100	0.960	0.974	0.979	0.983	0.987	0.989	0.991	0.991	0.985
	200	0.974	0.983	0.986	0.987	0.989	0.990	0.992	0.991	0.989
	500	0.981	0.984	0.987	0.988	0.988	0.990	0.990	0.990	0.988
	1000	0.985	0.986	0.987	0.988	0.989	0.989	0.990	0.990	0.988
	5000	0.989	0.988	0.989	0.989	0.990	0.991	0.990	0.989	0.991
6	100	0.950	0.969	0.979	0.983	0.985	0.988	0.991	0.990	0.983
	200	0.967	0.979	0.983	0.985	0.987	0.989	0.990	0.991	0.987
	500	0.978	0.984	0.987	0.989	0.990	0.991	0.991	0.991	0.986
	1000	0.982	0.985	0.987	0.987	0.988	0.989	0.989	0.989	0.988
	5000	0.989	0.988	0.989	0.989	0.989	0.990	0.992	0.992	0.990
10	200	0.940	0.964	0.976	0.980	0.984	0.988	0.992	0.991	0.982
	1000	0.975	0.982	0.984	0.987	0.988	0.989	0.989	0.989	0.984
	5000	0.986	0.988	0.988	0.989	0.989	0.990	0.990	0.990	0.990
	10000	0.987	0.989	0.989	0.989	0.990	0.991	0.991	0.990	0.990
20	1000	0.962	0.976	0.981	0.985	0.987	0.987	0.987	0.988	0.985
	5000	0.983	0.987	0.988	0.988	0.989	0.989	0.991	0.991	0.991
	10000	0.985	0.988	0.989	0.989	0.989	0.990	0.990	0.991	0.989
	100000	0.990	0.991	0.990	0.990	0.991	0.990	0.990	0.990	0.990
100	5000	0.963	0.979	0.983	0.987	0.988	0.988	0.989	0.990	0.989
	10000	0.977	0.983	0.986	0.988	0.988	0.989	0.990	0.991	0.990
	100000	0.989	0.991	0.990	0.991	0.990	0.990	0.990	0.990	0.991

Table A.12: Halfwidths (multiplied by $\sqrt{n}$) of two-sided 99% CIs of expectiles for the TAR(1) process with innovations $t(3)$ using the d -fold method with sample size n and expectile level α with 10,000 replications

d	n	$\alpha = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2	100	228.94	162.62	133.64	117.45	107.75	102.98	103.85	115.06	161.58
	200	252.38	173.72	140.86	122.37	111.12	105.05	104.45	114.25	160.63
	500	261.66	176.86	142.76	123.80	112.14	105.67	104.96	115.16	161.93
	1000	265.59	178.36	143.80	124.76	113.08	106.51	105.64	115.85	163.68
	5000	272.83	181.97	145.73	125.68	113.33	106.45	105.77	116.09	164.43
3	100	37.922	27.538	22.877	20.204	18.589	17.783	17.938	19.924	27.819
	200	42.200	29.305	23.857	20.813	18.968	17.989	17.992	19.850	28.067
	500	44.809	30.273	24.401	21.149	19.163	18.074	17.970	19.699	27.805
	1000	46.274	31.023	24.938	21.562	19.489	18.317	18.137	19.839	28.021
	5000	47.146	31.497	25.274	21.823	19.698	18.508	18.320	19.994	28.236
4	100	22.469	16.557	13.813	12.219	11.259	10.806	10.941	12.169	16.872
	200	25.414	17.825	14.540	12.684	11.559	10.966	10.975	12.115	17.125
	500	27.395	18.539	14.946	12.950	11.733	11.069	11.014	12.105	17.143
	1000	28.074	18.819	15.135	13.103	11.867	11.185	11.113	12.182	17.203
	5000	28.715	19.172	15.394	13.298	12.012	11.292	11.189	12.231	17.288
5	100	17.482	13.093	11.003	9.776	9.037	8.691	8.818	9.830	13.539
	200	20.118	14.268	11.689	10.219	9.327	8.863	8.889	9.848	13.927
	500	21.978	14.908	12.012	10.401	9.421	8.889	8.850	9.736	13.861
	1000	22.541	15.073	12.112	10.486	9.503	8.967	8.922	9.805	13.929
	5000	23.234	15.477	12.410	10.711	9.667	9.078	8.986	9.828	13.928
6	100	14.935	11.325	9.560	8.517	7.887	7.599	7.728	8.634	11.831
	200	17.527	12.521	10.282	8.999	8.221	7.823	7.862	8.740	12.384
	500	19.448	13.239	10.679	9.254	8.386	7.916	7.885	8.683	12.379
	1000	19.997	13.380	10.745	9.294	8.412	7.927	7.878	8.657	12.351
	5000	20.559	13.679	10.970	9.473	8.558	8.047	7.976	8.737	12.402
10	200	13.524	9.989	8.310	7.324	6.725	6.430	6.504	7.274	10.185
	1000	16.524	11.072	8.877	7.669	6.937	6.538	6.507	7.179	10.354
	5000	17.032	11.272	9.022	7.784	7.029	6.608	6.554	7.192	10.268
	10000	17.121	11.355	9.089	7.839	7.072	6.640	6.573	7.193	10.222
20	1000	14.433	9.837	7.899	6.823	6.171	5.822	5.817	6.479	9.501
	5000	15.448	10.135	8.077	6.955	6.275	5.897	5.851	6.434	9.272
	10000	15.443	10.153	8.103	6.981	6.297	5.916	5.866	6.440	9.222
	100000	15.469	10.242	8.181	7.042	6.343	5.949	5.887	6.445	9.153
100	5000	14.170	9.460	7.505	6.432	5.788	5.442	5.433	6.098	9.221
	10000	14.774	9.584	7.561	6.472	5.819	5.463	5.434	6.048	9.059
	100000	14.664	9.547	7.588	6.519	5.867	5.500	5.443	5.974	8.581

(A4) In our setting, $\mathbf{x}_i = 1$. Then $\Sigma_n = \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top = n$, so $\max_{i \leq n} \mathbf{x}_i^\top \Sigma_n^{-1/2} \mathbf{x}_i = n^{-1/2} = o(1)$.

Hence conditions (A1)-(A4) in Theorem 1 in Wu [2007b] are all satisfied. Let the shift process $\mathcal{F}_k = (\dots, \varepsilon_{k-1}, \varepsilon_k)$ and $\mathcal{F}'_k = (\dots, \varepsilon_{-1}, \varepsilon'_0, \varepsilon_1, \dots, \varepsilon_k)$. Then $\mathcal{F}'_k$ is a coupled version of $\mathcal{F}_k$ with ε_0 replaced by ε'_0 and $\mathcal{F}'_j = \mathcal{F}_j$ for $j < 0$. It remains to verify the SRD condition

$$\sum_{k=0}^{\infty} \sup_w \|\mathbb{E}[g_\alpha(X_k + w)|\mathcal{F}_0] - \mathbb{E}[g_\alpha(X'_k + w)|\mathcal{F}'_0]\| < \infty. \quad (\text{A.2})$$

We now apply the coupling argument to prove the above relation. For any w , we have

$$\begin{aligned} \|\mathbb{E}[g_\alpha(X_k + w)|\mathcal{F}_0] - \mathbb{E}[g_\alpha(X'_k + w)|\mathcal{F}'_{-1}]\| &= \|\mathbb{E}[g_\alpha(X_k + w)|\mathcal{F}_0] - \mathbb{E}[g_\alpha(X'_k + w)|\mathcal{F}_0]\| \\ &\leq \|g_\alpha(X_k + w) - g_\alpha(X'_k + w)\| \leq \|X_k - X'_k\| \end{aligned}$$

and similarly

$$\|\mathbb{E}[g_\alpha(X_k + w)|\mathcal{F}_0] - \mathbb{E}[g_\alpha(X'_k + w)|\mathcal{F}'_{-1}]\| \leq \|X_k - X'_k\|.$$

Therefore,

$$\|\mathbb{E}[g_\alpha(X_k + w)|\mathcal{F}_0] - \mathbb{E}[g_\alpha(X'_k + w)|\mathcal{F}'_0]\| \leq 2\|X_k - X'_k\|.$$

Thus condition (7) in Wu [2007b] is satisfied in view of (2.9), and

$$\sum_{k=0}^{\infty} \sup_w \|\mathbb{E}[g_\alpha(X_k + w)|\mathcal{F}_0] - \mathbb{E}[g_\alpha(X'_k + w)|\mathcal{F}'_0]\| \leq 2 \sum_{k=0}^{\infty} \delta_{k,2} < \infty.$$

By Theorem 1 in Wu [2007b], we obtain the Bahadur representation

$$\sqrt{n}(\hat{\theta}_{\alpha,1} - \theta_{\alpha})\varphi'(0) - \frac{1}{\sqrt{n}} \sum_{t=1}^n g_{\alpha}(X_t - \theta_{\alpha}) \xrightarrow{\mathbb{P}} 0, \quad (\text{A.3})$$

which is consistent with the result under i.i.d. assumption in Theorem 3 in Holzmam and Klar [2016]. For the general case with $d > 1$, due to stationarity, (A.3) implies that for all j ,

$$\sqrt{m}(\hat{\theta}_{\alpha,j} - \theta_{\alpha})\varphi'(0) - \frac{1}{\sqrt{m}} \sum_{t=1+(j-1)m}^{jm} g_{\alpha}(X_t - \theta_{\alpha}) \xrightarrow{\mathbb{P}} 0. \quad (\text{A.4})$$

We can derive the marginal asymptotic distributions for $\sqrt{m}(\hat{\theta}_{\alpha,j} - \theta_{\alpha})$, $j = 1, \dots, d$, in the view of (A.4). However, we still need to investigate the joint distribution of $\sqrt{m}(\hat{\theta}_{\alpha,1} - \theta_{\alpha}, \hat{\theta}_{\alpha,2} - \theta_{\alpha}, \dots, \hat{\theta}_{\alpha,d} - \theta_{\alpha})$. To this end, define the partial sum process S_u , $0 \leq u \leq n$, as

$$S_0 = 0, \quad S_k = \sum_{i=1}^k g_{\alpha}(X_i - \theta_{\alpha}), \quad k = 1, \dots, n,$$

$$S_u = S_k + (u - k)g_{\alpha}(X_{k+1} - \theta_{\alpha}), \quad k \leq u \leq k + 1, \quad k = 0, 1, \dots, n - 1.$$

Under condition (2.9), we have the invariance principle (Theorem 3 (i) in Wu [2005b], Theorem 3 (ii) in Wu [2011])

$$\left\{ \frac{S_{nu}}{\sqrt{n}}, 0 \leq u \leq 1 \right\} \Rightarrow \sigma_{\alpha\infty} \{ \mathbb{B}(u), 0 \leq u \leq 1 \}, \quad (\text{A.5})$$

where $\mathbb{B}(\cdot)$ is the standard Brownian motion and the long-run variance

$$\sigma_{\alpha\infty} = \sum_{t=-\infty}^{\infty} \text{Cov}\{g_{\alpha}(X_1 - \theta_{\alpha}), g_{\alpha}(X_t - \theta_{\alpha})\} < \infty.$$

Let $u = k/d$, $k = 1, \dots, d$, by the definition of Brownian motion, we have that $\mathbb{B}(k/d) -$

$\mathbb{B}((k-1)/d)$, $k = 1, \dots, d$, are i.i.d. $\mathbf{N}(0, 1/d)$. By (A.5),

$$\frac{1}{\sqrt{n}} \left(\sum_{t=1}^m g_\alpha(X_t - \theta_\alpha), \sum_{t=m+1}^{2m} g_\alpha(X_t - \theta_\alpha), \dots, \sum_{t=m(d-1)+1}^n g_\alpha(X_t - \theta_\alpha) \right) \Rightarrow \frac{\sigma_{\alpha\infty}}{\sqrt{d}} \mathbf{N}(0, \mathbf{I}_d). \quad (\text{A.6})$$

With (A.4) and (A.6), we have (2.10) by Slutsky's theorem. $\square$

A.2.2 Proof of Corollary 2.2.1.1

Proof. For fixed d , let $Y_{n,j} = \sqrt{m}(\hat{\theta}_{\alpha,j}(X_1, \dots, X_n) - \theta_\alpha)$ and $Y_n = (Y_{n,1}, \dots, Y_{n,d})^\top$. Let $Z = (Z_1, \dots, Z_d) \sim \mathbf{N}(0, \sigma_{\alpha\infty}^2/\tau_\alpha^2 \mathbf{I}_d)$. By Theorem 1, we have $Y_n \Rightarrow Z$. Let continuous function $h : \mathbb{R}^d \mapsto \mathbb{R}$ be

$$h(y) = h(y_1, \dots, y_d) = \frac{\sqrt{d}\bar{y}}{\sqrt{\sum_{j=1}^d (y_j - \bar{y})^2/(d-1)}},$$

where $\bar{y} = (y_1 + \dots + y_d)/d$. By the continuous mapping theorem, we have $h(Y_n) \Rightarrow h(Z)$, where

$$h(Y_n) = \frac{\sqrt{d}\bar{Y}_n}{\sqrt{\frac{1}{d-1} \sum_{j=1}^d (Y_{n,j} - \bar{Y}_n)^2}} = \frac{\frac{\sqrt{m}}{d} \sum_{j=1}^d (\hat{\theta}_{\alpha,j} - \theta_\alpha)}{\sqrt{\frac{m}{d-1} \sum_{j=1}^d (\hat{\theta}_{\alpha,j} - \bar{\theta}_{\alpha,d})^2}} = \frac{\sqrt{d}(\bar{\theta}_{\alpha,d} - \theta_\alpha)}{\hat{\sigma}_{\alpha,d}}$$

and it is well-known that $h(Z) \sim t_{d-1}$. $\square$

A.2.3 Proof of Theorem 2.2.2

We first present a lemma, which will be frequently referenced in the proof of Theorem 2.2.2.

Lemma A.2.1. (i) Under condition (i) of Theorem 2.2.2, for

$$M_n(\theta) = n^{-1/2} \sum_{i=1}^n (|X_i - \theta| - \mathbb{E}|X_i - \theta|), \quad (\text{A.7})$$

there exists a constant $C > 0$, such that for any $\phi \in \mathbb{R}$ and $0 < \zeta \leq 1$, the maximal inequality holds:

$$\| \max_{\phi \leq \theta \leq \phi + \zeta} |M_n(\theta) - M_n(\phi)| \|_p \leq C \zeta^{1-1/K}. \quad (\text{A.8})$$

(ii) Under condition (ii) of Theorem 2.2.2, we have

$$\| \max_{\phi \leq \theta \leq \phi + \zeta} |M_n(\theta) - M_n(\phi)| \| \leq C \zeta. \quad (\text{A.9})$$

Proof. Let $\theta, \theta' \in \mathbb{R}$. Recall $\mathcal{F}_t = (\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots)$ and define the projection operator

$$\mathcal{P}_t \cdot = \mathbb{E}(\cdot | \mathcal{F}_t) - \mathbb{E}(\cdot | \mathcal{F}_{t-1}). \quad (\text{A.10})$$

Note that $\|\mathcal{P}_0 |X_i - \theta|\|_p \leq \delta_{i,p}$ and $||x - \theta| - |x - \theta'|| \leq |\theta - \theta'|$. By (2.14), we have

$$\begin{aligned} \sum_{i=0}^{\infty} \|\mathcal{P}_0(|X_i - \theta| - |X_i - \theta'|)\|_p &\leq \sum_{i=0}^{\infty} \min(|\theta - \theta'|, 2\delta_{i,p}) \\ &\leq C_2 \min(\delta_{0,p}, |\theta - \theta'|^{1-1/K}) < \infty \end{aligned} \quad (\text{A.11})$$

in view of $\sum_{i=1}^{\infty} \min(b, i^{-K}) = O(b^{1-1/K})$. By Theorem 1(iii) in Wu [2007a],

$$\|M_n(\theta) - M_n(\theta')\|_p \leq c_p \sum_{i=0}^{\infty} \|\mathcal{P}_0(|X_i - \theta| - |X_i - \theta'|)\|_p \leq C_3 \min(1, |\theta - \theta'|^{1-1/K}), \quad (\text{A.12})$$

where the constant C_3 is independent of n . Let $d \in \mathbb{N}$, $\phi_j = \phi + \zeta j / 2^d$, $j = 0, \dots, 2^d$, and $Z_j = M(\phi_j) - M(\theta_{j-1})$. By the moment maximal inequality in Proposition 1(i) in Wu

[2007a] and (A.12),

$$\begin{aligned}
\left\| \max_{1 \leq j \leq 2^d} |M(\phi_j) - M(\phi)| \right\|_p &\leq \sum_{r=0}^d \left[\sum_{m=1}^{2^{d-r}} \|M(\phi_{2^r m}) - M(\phi_{2^r(m-1)})\|_p^p \right]^{1/p} \\
&\leq \sum_{r=0}^d [2^{d-r} C_3^p (2^r \zeta / 2^d)^{(1-1/K)p}]^{1/p} \\
&\leq C_3 \zeta^{1-1/K} \sum_{h=0}^{\infty} 2^{h(1/p+1/K-1)} = C_5 \zeta^{1-1/K}
\end{aligned}$$

since $1/p + 1/K - 1 < 0$. Notice that $\max_{1 \leq j \leq 2^d} |M(\phi_j) - M(\phi)|$ is decreasing in d and its limit as $d \rightarrow \infty$ is $\max_{\phi \leq \theta \leq \phi+\zeta} |M_n(\theta) - M_n(\phi)|$, then (A.8) follows.

(ii) We now prove (A.9). Recall the empirical distribution function

$$F_n(\theta) = n^{-1} \sum_{i=1}^n \mathbf{1}_{X_i \leq \theta}. \quad (\text{A.13})$$

Then for $\theta > \phi$, we have

$$n^{-1/2}(M_n(\theta) - M_n(\phi)) = 2 \int_{\phi}^{\theta} (F_n(u) - F(u)) du. \quad (\text{A.14})$$

Denote by κ the sum in (2.15). By Theorem 1(iii) in Wu [2007a], $\sup_u \sqrt{n} \|F_n(u) - F(u)\| \leq \kappa < \infty$. Thus

$$\begin{aligned}
\left\| \max_{\phi \leq \theta \leq \phi+\zeta} |M_n(\theta) - M_n(\phi)| \right\| &\leq 2\sqrt{n} \left\| \int_{\phi}^{\phi+\zeta} |F_n(u) - F(u)| du \right\| \\
&\leq 2\sqrt{n} \int_{\phi}^{\phi+\zeta} \|F_n(u) - F(u)\| du \leq 2\kappa\zeta,
\end{aligned}$$

completing the proof. □

Proof of Theorem 2.2.2. We first prove (2.16) under (i). We shall first show that

$$\sup\{\sqrt{n}|\hat{\theta}_\alpha - \theta_\alpha|, \alpha_1 \leq \alpha \leq \alpha_2\} = O_{\mathbb{P}}(1). \quad (\text{A.15})$$

To this end, let

$$Q_{\alpha,n}(u) = \frac{1}{n} \sum_{t=1}^n g_\alpha(X_t - u) \text{ and } Q_\alpha(u) = \mathbb{E}Q_{\alpha,n}(u). \quad (\text{A.16})$$

Since $F(\cdot)$ is continuous, for any $u_1 < u_2$, we have as $y \rightarrow 0$ that

$$\max_{\alpha_1 \leq \alpha \leq \alpha_2} \max_{u_1 \leq u \leq u_2} |Q_\alpha(u+y) - Q_\alpha(u) - yQ'_\alpha(u)| = o(y), \quad (\text{A.17})$$

where the derivative

$$Q'_\alpha(u) = -\alpha(1 - F(u)) + (\alpha - 1)F(u). \quad (\text{A.18})$$

Note that $Q'_\alpha(\theta_\alpha) = -\tau_\alpha$. Since $g_\alpha(x) = x/2 + (\alpha - 1/2)|x|$, we have

$$Q_{\alpha,n}(u) - Q_\alpha(u) = \bar{X}_n/2 - \mathbb{E}\left(\sum_{t=1}^n X_t\right)/2n + (\alpha - 1/2)M_n(u)/\sqrt{n}, \quad (\text{A.19})$$

By Lemma A.2.1(i),

$$\max_{\alpha_1 \leq \alpha \leq \alpha_2} \max_{u_1 \leq u \leq u_2} |Q_{\alpha,n}(u) - Q_\alpha(u)| = O_{\mathbb{P}}(n^{-1/2}). \quad (\text{A.20})$$

Then for any sequence r_n , $r_n/\sqrt{n} \rightarrow 0$, we have $Q_{\alpha,n}(\theta_\alpha - r_n/\sqrt{n}) > 0$ and $Q_{\alpha,n}(\theta_\alpha + r_n/\sqrt{n}) < 0$ uniformly over $\alpha_1 \leq \alpha \leq \alpha_2$ with probability going to 1. Since $Q_{\alpha,n}(\hat{\theta}_\alpha) = 0$, by the monotonicity of $Q_{\alpha,n}(\cdot)$, (A.15) holds.

Again by Lemma A.2.1, we have uniformly over $\alpha_1 \leq \alpha \leq \alpha_2$ that

$$Q_{\alpha,n}(\hat{\theta}_\alpha) - Q_\alpha(\hat{\theta}_\alpha) = Q_{\alpha,n}(\theta_\alpha) - Q_\alpha(\theta_\alpha) + n^{-1/2}O_{\mathbb{P}}((n^{-1/2})^{1-1/K}) \quad (\text{A.21})$$

in view of (A.15). Note that $Q_{\alpha,n}(\hat{\theta}_\alpha) = 0$, $Q_\alpha(\theta_\alpha) = 0$, by the uniform Taylor expansion (A.17), we have

$$\sqrt{n}(\hat{\theta}_\alpha - \theta_\alpha) = \frac{\sqrt{n}Q_{\alpha,n}(\theta_\alpha)}{Q'_\alpha(\theta_\alpha)} + O_{\mathbb{P}}((n^{-1/2})^{1-1/K}) \quad (\text{A.22})$$

uniformly over $\alpha_1 \leq \alpha \leq \alpha_2$. To prove the weak convergence (2.16), we need to verify the finite dimensional convergence and the tightness (cf. Billingsley [2013]). The finite dimensional convergence follows from the Bahadur representation (A.3) and the central limit theorem of $\sqrt{n}Q_{\alpha,n}(\theta_\alpha)$ via the Cramér-Wold device. We now prove the tightness of $\{\sqrt{n}Q_{\alpha,n}(\theta_\alpha), \alpha_1 \leq \alpha \leq \alpha_2\}$, which by the continuous mapping theorem implies the tightness of $\{\sqrt{n}Q_{\alpha,n}(\theta_\alpha)/Q'_\alpha(\theta_\alpha), \alpha_1 \leq \alpha \leq \alpha_2\}$. Let

$$R_{\alpha,n}(\theta) = \sqrt{n}[Q_{\alpha,n}(\theta) - Q_\alpha(\theta)]. \quad (\text{A.23})$$

For $\alpha, \alpha' \in [\alpha_1, \alpha_2]$, we have

$$|R_{\alpha,n}(\theta_\alpha) - R_{\alpha',n}(\theta_{\alpha'})| \leq |R_{\alpha,n}(\theta_\alpha) - R_{\alpha',n}(\theta_\alpha)| + |R_{\alpha',n}(\theta_\alpha) - R_{\alpha',n}(\theta_{\alpha'})|. \quad (\text{A.24})$$

Since $2x^+ = x + |x|$ and $2x^- = x - |x|$, recall $S_n = \sum_{i=1}^n X_i$, by (A.19),

$$|R_{\alpha,n}(\theta_\alpha) - R_{\alpha',n}(\theta_\alpha)| \leq |\alpha - \alpha'| |M_n(\theta_\alpha)|, \quad (\text{A.25})$$

and

$$|R_{\alpha',n}(\theta_\alpha) - R_{\alpha',n}(\theta_{\alpha'})| \leq |M_n(\theta_\alpha) - M_n(\theta_{\alpha'})|. \quad (\text{A.26})$$

Notice that θ_α is uniformly continuous over $\alpha_1 \leq \alpha \leq \alpha_2$. Let $u_1 = \theta_{\alpha_1}$ and $u_2 = \theta_{\alpha_2}$. By Lemma A.2.1,

$$\mathbb{P} \left[\sup_{u, u' \in [u_1, u_2], |u - u'| \leq \zeta} |M_n(u) - M_n(u')| \geq \epsilon \right] \leq C \frac{u_2 - u_1}{\zeta \epsilon^p} \zeta^{(1-1/K)p}, \quad (\text{A.27})$$

which converges to 0 as $\zeta \rightarrow 0$ since $(1 - 1/K)p > 1$. Thus the process $\{M_n(u), u_1 \leq u \leq u_2\}$ is tight, which implies the tightness of $\{M_n(\theta_\alpha), \alpha_1 \leq \alpha \leq \alpha_2\}$. Observe that Lemma A.2.1 implies $\|\max_{\phi \leq \theta \leq \phi+1} |M_n(\theta)|\|_p \leq C$. Thus $\|\max_{u_1 \leq u \leq u_2} |M_n(\theta)|\|_p \leq C_1$ since u_1, u_2 are finite. By (A.24)-(A.26), the process $\{\sqrt{n}Q_{\alpha,n}(\theta_\alpha), \alpha_1 \leq \alpha \leq \alpha_2\}$ is tight, completing the proof of (2.16) under (i).

We shall now prove (2.16) under (ii) using the same argument as in (i). By Lemma A.2.1(ii), we have (A.20), which implies (A.15). The error term $O_{\mathbb{P}}((n^{-1/2})^{1-1/K})$ in (A.22) now becomes the sharper one $O_{\mathbb{P}}(n^{-1/2})$, and hence the finite dimensional convergence follows. The right hand side of (A.27) becomes $C(\zeta\epsilon^2)^{-1}(u_2 - u_1)\zeta^2 = C(u_2 - u_1)\zeta/\epsilon^2$, still ensuring the tightness. Thus the weak convergence (2.16) similarly follows. $\square$

A.2.4 Proof of Theorem 2.2.4

Proof of Theorem 2.2.4. Let $m(\theta) = \mathbb{E}|X_i - \theta|$, $m'(\theta) = 2F(\theta) - 1$, and

$$N_n(\theta) = \sum_{i=1}^n R_i(\theta), \text{ where } R_i(\theta) = |X_i - \theta| - m(\theta) - m'(\theta)X_i. \quad (\text{A.28})$$

Then $R'_i(\theta) = 2 \times \mathbf{1}_{X_i \leq \theta} - 2F(\theta) - 2f(\theta)X_i$. Let $A_n(j) = \sum_{t=n}^{\infty} |a_t|^j$, $\phi_n = |a_{n-1}|[|a_{n-1}| + A_n(p)^{1/2} + A_n(2)^{1/2}] = O(|a_{n-1}|A_n(2)^{1/2})$, $\Theta_n = \sum_{t=1}^n \phi_t$. By Corollary 1 in Wu [2006], under the assumption $\mathbb{E}|\varepsilon_i|^p < \infty, p > 2$, and $\sup_x (f_\varepsilon(x) + |f'_\varepsilon(x)|) < \infty$, we have

$$\sup_{\theta} (\|N(\theta)\|^2 + \|N'(\theta)\|^2) = O(\Xi_n) \text{ where } \Xi_n = n\Theta_n + \sum_{i=1}^{\infty} (\Theta_{n+i} - \Theta_i)^2, \quad (\text{A.29})$$

where $\Xi_n = O(n)$ if $\beta > 3/4$, $\Xi_n = O(n^{4-4\beta}\ell^4(n))$ if $1/2 < \beta < 3/4$ and $\Xi_n = O(n[\sum_{t=1}^n \ell^2(t)/t]^2)$ if $\beta = 3/4$. For all those three cases we have $\Xi_n = o(n^{3-2\beta}\ell^2(n))$. Hence for any $\theta_1 < \theta_2$,

$$\sup_{\theta \in [\theta_1, \theta_2]} |N(\theta)| \leq |N(\theta_1)| + \sup_{\theta \in [\theta_1, \theta_2]} |N(\theta) - N(\theta_1)| \leq |N(\theta_1)| + \int_{\theta_1}^{\theta_2} |N'(\theta)| d\theta, \quad (\text{A.30})$$

which by (A.29) implies

$$\| \sup_{\theta \in [\theta_1, \theta_2]} |N(\theta)| \| = O(\Xi_n^{1/2}) = o(n^{3/2-\beta}\ell(n)). \quad (\text{A.31})$$

Note that $g_\alpha(x) = x/2 + (\alpha - 1/2)|x|$. Let $Q_\alpha(u) = \mathbb{E}g_\alpha(X_i - u)$. Thus $Q'_\alpha(u) = (\alpha - 1/2)(2F(u) - 1) - 1/2$. By (A.31),

$$\sup_{\alpha \in [\alpha_1, \alpha_2]} \sup_{\theta \in [\theta_1, \theta_2]} \left| \sum_{i=1}^n [g_\alpha(X_i - \theta) - Q_\alpha(\theta) - Q'_\alpha(\theta)X_i] \right| = o(n^{3/2-\beta}\ell(n)). \quad (\text{A.32})$$

Thus for any sequence $L_n \rightarrow \infty$, we have

$$\inf_{\alpha \in [\alpha_1, \alpha_2]} \sum_{i=1}^n [g_\alpha(X_i - \theta_\alpha - L_n n^{1/2-\beta}\ell(n))] > 0 \text{ and } \sup_{\alpha \in [\alpha_1, \alpha_2]} \sum_{i=1}^n [g_\alpha(X_i + \theta_\alpha - L_n n^{1/2-\beta}\ell(n))] < 0.$$

Since the rate of $L_n \rightarrow \infty$ can be arbitrarily, we have

$$\sup_{\alpha \in [\alpha_1, \alpha_2]} |\hat{\theta}_\alpha - \theta_\alpha| = O_{\mathbb{P}}(n^{1/2-\beta}\ell(n)). \quad (\text{A.33})$$

Since $F(\cdot)$ has continuous derivative, (A.33) implies that

$$\sup_{\alpha \in [\alpha_1, \alpha_2]} |Q'_\alpha(\hat{\theta}_\alpha) - Q'_\alpha(\theta_\alpha)| = O_{\mathbb{P}}(n^{1/2-\beta}\ell(n)). \quad (\text{A.34})$$

By (A.32) and (A.33), we obtain

$$\sup_{\alpha \in [\alpha_1, \alpha_2]} \left| \sum_{i=1}^n [g_\alpha(X_i - \hat{\theta}_\alpha) - Q_\alpha(\hat{\theta}_\alpha) - Q'_\alpha(\hat{\theta}_\alpha)X_i] \right| = o(n^{3/2-\beta}\ell(n)). \quad (\text{A.35})$$

Note that $\sum_{i=1}^n g_\alpha(X_i - \hat{\theta}_\alpha) = 0$, $Q_\alpha(\theta_\alpha) = 0$ and $Q_\alpha(\hat{\theta}_\alpha) = (\hat{\theta}_\alpha - \theta_\alpha)Q'_\alpha(\theta_\alpha) + o_{\mathbb{P}}(n^{1/2-\beta}\ell(n))$, we obtain (2.22), which implies (2.23) in view of (2.21). $\square$

A.2.5 Proof of Lemma 2.3.3

Proof. Let

$$\begin{aligned} M_n(\theta) &= \sum_{i=1}^n \{(X_i - \theta)\mathbb{1}_{X_i \leq \theta} - \mathbb{E}[(X_i - \theta)\mathbb{1}_{X_i \leq \theta}]\}, \\ N_n(\theta) &= \sum_{i=1}^n \{\theta\mathbb{1}_{X_i \leq \theta} - \mathbb{E}[\theta\mathbb{1}_{X_i \leq \theta}]\}. \end{aligned}$$

Then $H_n(\theta) = M_n(\theta) + N_n(\theta)$. By Fubini's theorem, we have

$$\begin{aligned} M_n(\theta) - M_n(\theta_0) &= \int_{\theta_0}^{\theta} \sum_{i=1}^n \mathbb{1}_{X_i \leq u} du - \mathbb{E} \left[\int_{\theta_0}^{\theta} \sum_{i=1}^n \mathbb{1}_{X_i \leq u} du \right] \\ &= \int_{\theta_0}^{\theta} n[F_n(u) - F(u)] du. \end{aligned}$$

Therefore, by Lemma 2.3.2, for any $\iota > 0$, we have

$$\begin{aligned} \sup_{|\theta - \theta_0| \leq b_n} |M_n(\theta) - M_n(\theta_0)| &\leq n \int_{\theta_0 - b_n}^{\theta_0 + b_n} |F_n(u) - F(u)| du \\ &= O_{\text{a.s.}}(n^{1/2}b_n(\log n)^{3/2}(\log \log n)^{1/2+\iota/4}) \end{aligned} \quad (\text{A.36})$$

On the other hand, observe that

$$N_n(\theta) - N_n(\theta_0) = \theta n[F_n(\theta) - F(\theta) + F_n(\theta_0) - F(\theta_0)] + n(\theta - \theta_0)[F_n(\theta_0) - F(\theta_0)].$$

Again, by Lemma 2.3.2, we have

$$\begin{aligned} \sup_{|\theta - \theta_0| \leq b_n} |N_n(\theta) - N_n(\theta_0)| &\leq n(|\theta_0| + b_n) \sup_{|\theta - \theta_0| \leq b_n} |F_n(\theta) - F(\theta) + F_n(\theta_0) - F(\theta_0)| \\ &\quad + nb_n |F_n(\theta_0) - F(\theta_0)| \\ &= O_{\text{a.s.}}(|\theta_0| n^{1/4} (\log n)^{q'}) \end{aligned} \tag{A.37}$$

for some $q' > 9/4$. Combining (A.36) and (A.37) completes the proof. $\square$

A.2.6 Proof of Theorem 2.3.4

Proof. By Lemma 2.3.1 and 2.3.3, we have

$$|H_n(\hat{q}_\alpha) - H_n(q_\alpha)| = O_{\text{a.s.}}(n^{1/4} (\log n)^{q'})$$

which indicates

$$-\alpha n \widehat{\text{ES}}_\alpha = n(g(\hat{q}_\alpha) - g(q_\alpha)) + \sum_{i=1}^n X_i \mathbb{1}_{X_i \leq q_\alpha} + O_{\text{a.s.}}(n^{1/4} (\log n)^{q'}).$$

By Taylor expansion and Lemma 2.3.1 on $g(\cdot)$, we can further obtain that

$$\begin{aligned}
-\alpha n \widehat{\text{ES}}_\alpha &= ng'(q_\alpha)(\hat{q}_\alpha - q_\alpha) + \sum_i X_i \mathbb{1}_{X_i \leq q_\alpha} + O_{\text{a.s.}}(n^{1/4}(\log n)^{q'}) \\
&= nq_\alpha(\alpha - F_n(q_\alpha)) + \sum_i X_i \mathbb{1}_{X_i \leq q_\alpha} + O_{\text{a.s.}}(n^{1/4}(\log n)^{q'}) \\
&= n\alpha q_\alpha + \sum_i h_\alpha(X_i) + O_{\text{a.s.}}(n^{1/4}(\log n)^{q'})
\end{aligned}$$

By CLT on $h_\alpha(X_i)$ and Slutsky's theorem, we have

$$\sqrt{n}(-\alpha \widehat{\text{ES}}_\alpha - \alpha q_\alpha) \xrightarrow{d} N(\mathbb{E}h_\alpha(X_1), \sigma_\infty^2)$$

where $\mathbb{E}h_\alpha(X_1) = -\alpha \text{ES}_\alpha - q_\alpha F(q_\alpha)$, which completes the proof. $\square$

A.2.7 Proof of Theorem 2.3.5

Proof. By Lemma 2.3.3, for $-\infty < \theta_1 \leq \theta_0 \leq \theta_2 < \infty$, we have

$$\sup_{\theta_1 \leq \theta_0 \leq \theta_2} \sup_{|\theta - \theta_0| \leq b_n} |H_n(\theta) - H_n(\theta_0)| = O_{\text{a.s.}}(\max(|\theta_1|, |\theta_2|)n^{1/4}(\log n)^{q'}),$$

which indicates we have uniformly over $\alpha_1 \leq \alpha \leq \alpha_2$ that

$$H_n(\hat{q}_\alpha) = H_n(q_\alpha) + O_{\text{a.s.}}(n^{1/4}(\log n)^{q'}).$$

Since $f \in \mathcal{C}^1$, by the uniform Taylor expansion on $g(\theta)$, we have that

$$\sup_{\theta_1 \leq \theta \leq \theta_2} |g(\theta + u) - g(\theta) - ug'(\theta)| = O(u^2)$$

as $u \rightarrow 0$. Then we have uniformly over $\alpha_1 \leq \alpha \leq \alpha_2$ that

$$\begin{aligned}\sqrt{n}\alpha \left[\widehat{\text{ES}}_\alpha - \text{ES}_\alpha + q_\alpha - \frac{q_\alpha}{\alpha} F(q_\alpha) \right] &= \frac{\sum_{i=1}^n [h_\alpha(X_i) - \mathbb{E}h_\alpha(X_i)]}{\sqrt{n}} + O_{\text{a.s.}}(n^{-3/4}(\log n)^{q'}) \\ &= \frac{M_n(q_\alpha)}{\sqrt{n}} + O_{\text{a.s.}}(n^{-3/4}(\log n)^{q'})\end{aligned}\tag{A.38}$$

The finite dimensional convergence of (A.38) can be verified based on Theorem 2.3.4 via Cramer-Wold theorem. For the tightness of $\{M_n(q_\alpha)/\sqrt{n}, \alpha_1 \leq \alpha \leq \alpha_2\}$, by (A.36), we have that for any $\delta > 0$ and $q > 1$

$$\mathbb{P} \left[\sup_{\alpha_1 \leq \alpha, \alpha' \leq \alpha_2} n^{-1/2} |M_n(q_\alpha) - M_n(q_{\alpha'})| > \delta \right] \leq C \frac{q_{\alpha_2} - q_{\alpha_1}}{b_n \delta^q} b_n^q (\log n)^{3q/2} (\log \log n)^{3q/4}$$

which converges to 0 as $b_n \rightarrow 0$. Thus, $\{M_n(q_\alpha)/\sqrt{n}, \alpha_1 \leq \alpha \leq \alpha_2\}$ is tight, which completes proving the weak convergence. $\square$

A.2.8 Proof of Theorem 2.3.6

Proof. Following Wu [2006], in our setting, we have $K(X) = X\mathbb{1}_{X \leq \theta} - \mathbb{E}[X\mathbb{1}_{X \leq \theta}]$. Let $K_\infty(w) = \mathbb{E}[K(w + X_n)]$. Then we have

$$\kappa_1 = K'_\infty(0) = \int_{-\infty}^{\theta} x f'(x) dx, \quad \kappa_2 = \int_{-\infty}^{\theta} x f''(x) dx.$$

Applying the Theorem 1 in Wu [2006], we can directly derive the results. $\square$

APPENDIX B

APPENDIX : SUPPLEMENTARY MATERIALS FOR CHAPTER

3

B.1 Factor pricing model

A factor model has been one of the most successful asset pricing models in finance. Given that related asset prices tend to move together in general, the principal component of these assets can explain a large share of the variance under a suitably constructed factor model. The class of these models include the single-factor capital asset pricing model (*CAPM*) and the Fama-French three factor model, etc. The methodology developed in this paper can be used to check whether the *pricing errors* in these factor models are indeed zero or not (Lewellen and Nagel [2006]). To this end, we consider the following generalization of a single-factor asset pricing model:

$$R_t = \alpha(\mathbf{Z}_t) + \beta(\mathbf{Z}_t)R_t^m + \sigma(\mathbf{Z}_t)\varepsilon_t \quad (\text{B.1})$$

where R_t is the excess return of momentum portfolio (Jegadeesh and Titman [1993]) and R_t^m is the excess return on the value-weighted market index portfolio at t , respectively. Here we let $\alpha(\cdot)$, the *pricing error* of the factor model, depend on $\mathbf{Z}_t$, a *vector* of state variables. The model pricing error in (B.1) is likely to be time-varying and its variation is related to the vector $\mathbf{Z}_t$ that is known to predict the excess portfolio return. To this end, we let $\mathbf{Z}_t$ be a bi-variate random vector of:

- (i) market portfolio's dividend yield
- (ii) de-trended T-bill yield (i.e. the lagged one-month T-bill yield less its 12-month moving average).

The key issue here is whether the pricing error of the factor model (B.1) is zero in $\mathbf{Z}_t$.

There has been a vast literature on testing the hypothesis of a zero pricing error for asset pricing factor models (Harvey [1989]; Jagannathan and Wang [1996]; Lewellen and Nagel [2006]; Ang and Kristensen [2012]; Engelberg et al. [2018]). To our best knowledge, however, the current study is the first one to model the pricing error of a single-factor *CAPM* as a multivariate unknown function of relevant state variables and to show how to perform simultaneous inference of the zero-pricing-error hypothesis under such a framework. Through the simultaneous inference proposed here, we not only get to test the hypothesis, but also get to better understand the relationship between the pricing error and the variables that have been known in the literature to predict the portfolio returns, such as the dividend yield or the T-bill yield. To this end, we consider the following generalized factor model with a *bivariate* pricing error:

$$R_t = \alpha(DY_{t-1}, TY_{t-1}) + \beta(DY_{t-1}, TY_{t-1})R_t^m + \sigma(DY_{t-1}, TY_{t-1})\varepsilon_t, \quad (\text{B.2})$$

where DY_t and TY_t represent the dividend yield and de-trended T-bill yield, respectively. Here (B.2) is used to test the validity of the single-factor *CAPM*, where the pricing error $\alpha(\cdot, \cdot)$ is an unknown bi-variate function of the dividend yield and the T-bill yield. Given that (B.2) is a special case of (3.1), our methodology readily applies. By constructing a *SCR* for the unknown $\alpha(\cdot, \cdot)$, one can conduct simultaneous inference for the zero-pricing-error hypothesis for the factor model in (B.2).

The U.S. monthly data for the momentum portfolio, the market excess returns and the two yield variables during JUL 1963–MAR 2019 are obtained from the Kenneth French’s data library (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). The estimation results are shown by Figs.B.1 and B.2.

Fig.B.1 provides the estimates of $\alpha(\cdot, \cdot)$ and $\beta(\cdot, \cdot)$ in terms of the dividend yield and the T-bill yield in a three-dimensional way. The figure clearly shows the *highly nonlinear* nature of the underlying relationship. Fig.B.2 represents the two-dimensional relationship

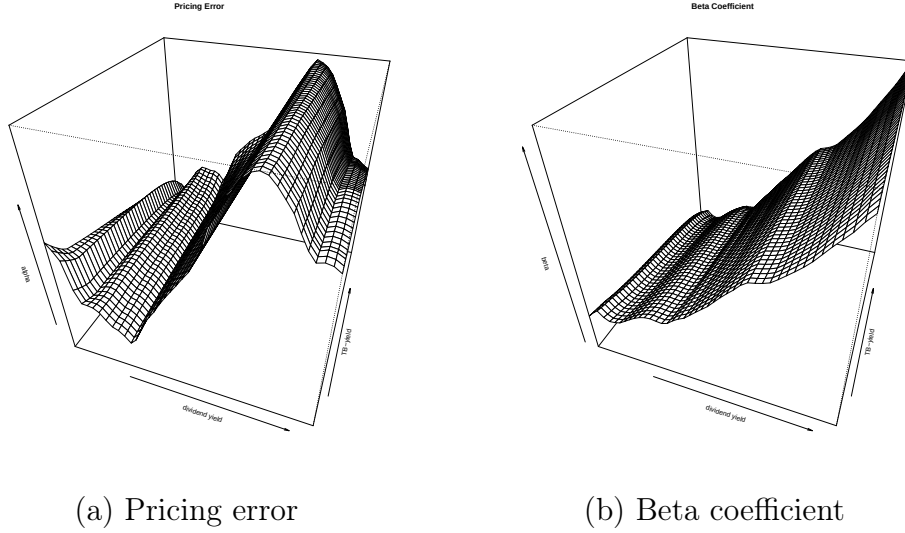
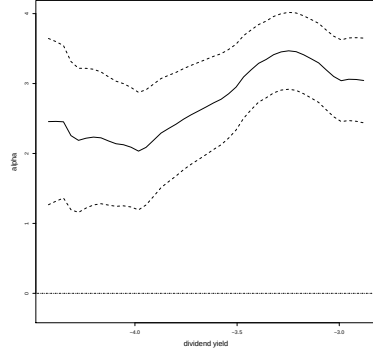
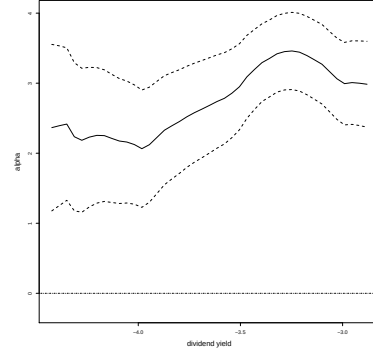


Figure B.1: Pricing error and the beta coefficient of a *CAPM* in the market portfolio's dividend yield and the de-trended T-bill yield; The bandwidth is obtained through the generalized cross-validation (*GCV*).

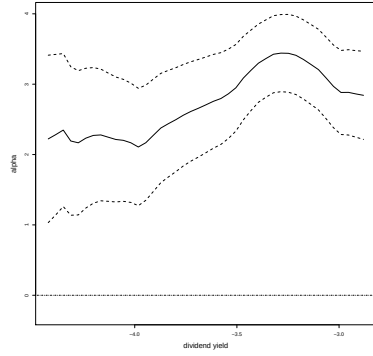
between the pricing error $\alpha(\cdot, \cdot)$ and the market portfolio's dividend yield when the T-bill yield is fixed at a certain percentile. Each panel in Fig.B.2 also includes the 95% *SCR* of the corresponding function estimate to conduct simultaneous inference of the zero-pricing-error hypothesis. To accept the null hypothesis of zero-pricing error at a 5% level, the horizontal line fixed at zero must be entirely contained by *all* *SCRs* in Fig.B.2. Clearly, Panels (a)–(c) in Fig.B.2 fail to contain the horizontal line entirely, and the null hypothesis thus gets rejected at a 5% level for the factor model in (B.2). The main reason for the rejection is because the estimated nonlinear pricing error with its relatively tight confidence band stays substantially higher than the null value of zero.



(a) 25th-percentile of T-bill yield



(b) 50th-percentile of T-bill yield



(c) 75th-percentile of T-bill yield

Figure B.2: The *solid* curve is the local-linear regression estimate of the pricing error $\alpha(\cdot)$. The *dashed* band is the 95% *SCR* of $\alpha(\cdot)$ and the dotted horizontal line is the null $H_0 : \alpha(\cdot) = 0$. The bandwidth is chosen by *GCV*.

B.2 Technical Proofs

B.2.1 Lemmas

Lemma B.2.1 (Bonnerjee et al. [2024]). *Let $X_1, \dots, X_n \in \mathbb{R}^p$, $p \geq 2$, be independent mean zero vectors such that $\mathbb{E}[|X_i|_\infty^q] < \infty$ for some $q > 2$. Let $\eta_1, \dots, \eta_n \in \mathbb{R}^p$ be independent mean zero Gaussian vectors such that for $1 \leq i \leq n$, $\text{Cov}(\eta_i) = \text{Cov}(X_i)$. Suppose $\sigma_j = \|\sum_{i=1}^n X_{ij}\| > 0$ for all $j \leq p$. Then*

$$\sup_{B \in \mathcal{A}^p} \left| \mathbb{P} \left(\sum_{i=1}^n X_i \in B \right) - \mathbb{P} \left(\sum_{i=1}^n \eta_i \in B \right) \right| \leq c_q (\log np)^{\frac{5q-6}{4q-2}} L_q^{\frac{1}{2q-1}}$$

where $L_q = \sum_{i=1}^n \mathbb{E}[|Y_i|_\infty^q]$ and c_q is a constant depending solely on q .

Proof of Proposition 3.2.1. Applying Lemma B.2.1 with $X_i = \varepsilon_i v_i$ and $\eta_i = \phi_i v_i$, we have

$$L_q = \sum_{i=1}^n \mathbb{E} \left[\max_{1 \leq j \leq p} \left| \frac{X_i v_{ij}}{\sigma_j} \right|^q \right] = m_q \sum_{i=1}^n \max_{j \leq p} \left| \frac{v_{ij}}{\sigma_j} \right|^q.$$

Thus, Proposition 3.2.1 follows from Lemma B.2.1. □

Lemma B.2.2. *Let*

$$f_n(z) = \frac{1}{nb^d} \sum_{i=1}^n X_i X_i^\top K \left(\frac{z - Z_i}{b} \right). \quad (\text{B.3})$$

Under the conditions of Theorem 3.2.1, we have

$$\mathbb{P} \left(\sup_{z \in \mathcal{D}} |f_n(z) - \mathbb{E} f_n(z)|_\infty > C n^{1/q-1/2} b^{-d/2} \log n \right) = o(1), \quad (\text{B.4})$$

where C is a constant independent of n and b .

Proof. Throughout the proof, we use C to denote a generic constant, which may vary from

line to line. Let

$$\begin{pmatrix} X_i^* \\ Z_i^* \end{pmatrix} = G(\dots, \eta'_{i,i-m-1}, \eta'_{i,i-m}, \eta_{i-m+1}, \dots, \eta_i) \quad (\text{B.5})$$

where $\{\eta'_{i,j}, j \in \mathbb{Z}\}$ is an i.i.d. copy of $\{\eta_j, j \in \mathbb{Z}\}$ for $i = 1, \dots, n$, and define the coupled version of $f_n(z)$ as

$$f_n^*(z) = \frac{1}{nb^d} \sum_{i=1}^n X_i^* X_i^{*\top} K\left(\frac{z - Z_i^*}{b}\right).$$

Comparing (B.5) and (3.10), we can see X_i^*, Z_i^* also satisfy the condition 3.2.2 and $\mathbb{E}f_n^*(z) = \mathbb{E}f_n(z)$. Then, firstly, we notice that

$$\begin{aligned} & \mathbb{P}\left(\sup_z |f_n(z) - \mathbb{E}f_n(z)|_\infty > t\right) \\ & \leq \mathbb{P}\left(\sup_z |f_n(z) - \mathbb{E}f_n(z)|_\infty > t, \max_i |X_i|_\infty \leq n^\tau, \max_i |X_i^*|_\infty \leq n^\tau\right) \\ & \quad + \mathbb{P}\left(\sup_z |f_n(z) - \mathbb{E}f_n(z)|_\infty > t, \max_i |X_i|_\infty > n^\tau \text{ or } \max_i |X_i^*|_\infty > n^\tau\right), \end{aligned} \quad (\text{B.6})$$

where $\tau = 1/q$. By Markov's inequality, the second term on the right hand side of (B.6) can be bounded as

$$\begin{aligned} & \mathbb{P}\left(\sup_z |f_n(z) - \mathbb{E}f_n(z)|_\infty > t, \max_i |X_i|_\infty > n^\tau \text{ or } \max_i |X_i^*|_\infty > n^\tau\right) \\ & \leq \frac{n}{n^{\tau q}} \mathbb{E}\left[|X_i|_\infty^q \mathbb{1}\{|X_i|_\infty \geq n^\tau\}\right] + \frac{n}{n^{\tau q}} \mathbb{E}\left[|X_i^*|_\infty^q \mathbb{1}\{|X_i^*|_\infty \geq n^\tau\}\right] = o(1). \end{aligned} \quad (\text{B.7})$$

Then, decompose $|f_n(z) - \mathbb{E}f_n(z)|_\infty$ into two parts

$$|f_n(z) - \mathbb{E}f_n(z)|_\infty \leq |f_n(z) - f_n^*(z)|_\infty + |f_n^*(z) - \mathbb{E}f_n^*(z)|_\infty. \quad (\text{B.8})$$

For the first part of (B.8), under the conditions $|X_i|_\infty \leq n^\tau$ and $|X_i^*|_\infty \leq n^\tau$, we have

$$\begin{aligned}
& \sup_z |f_n(z) - f_n^*(z)|_\infty \\
&= \sup_z \max_{j,k} \left| \frac{1}{nb^d} \sum_{i=1}^n \left(X_{ij} X_{ik} K\left(\frac{z - Z_i}{b}\right) - X_{ij}^* X_{ik}^* K\left(\frac{z - Z_i^*}{b}\right) \right) \right| \\
&\leq \frac{Ln^{2\tau-1}}{b^{1+d}} \sum_{i=1}^n \left| \begin{pmatrix} X_i^* \\ Z_i^* \end{pmatrix} - \begin{pmatrix} X_i \\ Z_i \end{pmatrix} \right|_\infty,
\end{aligned}$$

where the constant L is independent of n and b . By Condition 3.2.2 and the above inequality, we have the following bound

$$\begin{aligned}
& \mathbb{P} \left(\sup_z |f_n(z) - f_n^*(z)|_\infty > t, \max_i |X_i|_\infty \leq n^\tau, \max_i |X_i^*|_\infty \leq n^\tau \right) \\
&\leq \mathbb{P} \left(\frac{L}{b^{1+d}} n^{2\tau-1} \sum_{i=1}^n \left| \begin{pmatrix} X_i^* \\ Z_i^* \end{pmatrix} - \begin{pmatrix} X_i \\ Z_i \end{pmatrix} \right|_\infty > t \right) \\
&\leq \sum_{i=1}^n \mathbb{P} \left(\frac{L}{b^{1+d}} n^{2\tau-1} \left| \begin{pmatrix} X_i^* \\ Z_i^* \end{pmatrix} - \begin{pmatrix} X_i \\ Z_i \end{pmatrix} \right|_\infty > \frac{t}{n} \right) \\
&\leq L^\theta t^{-\theta} b^{-\theta(1+d)} n^{2\theta\tau+1} \mathbb{E} \left| \begin{pmatrix} X_i^* \\ Z_i^* \end{pmatrix} - \begin{pmatrix} X_i \\ Z_i \end{pmatrix} \right|_\infty^\theta \\
&\leq Ct^{-\theta} b^{-\theta(1+d)} n^{2\theta\tau+1} \rho^m.
\end{aligned} \tag{B.9}$$

Let $t = n^{\tau-1/2} b^{-d/2} \log n$. Then the bound (B.9) is $o(1/n)$ by choosing $m = \lceil c_1 \log n \rceil$, where c_1 is a large enough constant such that

$$n^{\tau\theta+\theta/2+1+c_1 \log \rho} b^{-\theta-d\theta/2} (\log n)^{-\theta} = o(n^{-1}).$$

For the second part of (B.8), we partition $[n] = \{1, \dots, n\}$ into m groups

$$[n] = \bigcup_{i=1}^m I_i = \bigcup_{i=1}^m \{i + jm : j \in \mathbb{Z}, 0 \leq i + jm \leq n\}.$$

By the definition of $f_n^\star(z)$ and $X_l^\star = (X_{l1}^\star, \dots, X_{lp}^\star)^\top$, we have

$$\begin{aligned} & \sup_z |f_n^\star(z) - \mathbb{E}f_n^\star(z)|_\infty \\ &= \sup_z \max_{j,k} \left| \frac{1}{nb^d} \sum_{i=1}^m \sum_{l \in I_i} \left(X_{lj}^\star X_{lk}^\star K\left(\frac{z - Z_l^\star}{b}\right) - \mathbb{E} \left[X_{lj}^\star X_{lk}^\star K\left(\frac{z - Z_l^\star}{b}\right) \right] \right) \right|. \end{aligned}$$

Thus, by the union bound,

$$\begin{aligned} & \mathbb{P} \left(\sup_z |f_n^\star(z) - \mathbb{E}f_n^\star(z)|_\infty > t, \max_i |X_i|_\infty \leq n^\tau, \max_i |X_i^\star|_\infty \leq n^\tau \right) \\ & \leq \sum_{j,k} \sum_{i=1}^m \mathbb{P} \left(\sup_z \left| f_{i,jk}^\diamond(z) - \mathbb{E}f_{i,jk}^\diamond(z) \right| > \frac{t}{m}, \max_i |X_i^\star|_\infty \leq n^\tau \right), \end{aligned} \quad (\text{B.10})$$

where

$$f_{i,jk}^\diamond(z) = \frac{1}{nb^d} \sum_{l \in I_i} X_{lj}^\star X_{lk}^\star K\left(\frac{z - Z_l^\star}{b}\right).$$

To bound (B.10), we partition the compact domain $\mathcal{D}$ into N subdomains $\{\mathcal{D}_u, u = 1, \dots, N\}$ such that for each subdomain $\mathcal{D}_u$, $|z - z'|_\infty \leq C/N$, $\forall z, z' \in \mathcal{D}_u$ with some constant C independent of n and b . Let z_u be an arbitrary point of $\mathcal{D}_u$. Under the condition $\max_i |X_i^\star|_\infty \leq n^\tau$, we have:

$$\begin{aligned} & |f_{i,jk}^\diamond(z) - f_{i,jk}^\diamond(z')| \\ & \leq \frac{1}{nb^d} \sum_{l \in I_i} \left| X_{lj}^\star X_{lk}^\star \right| \left| K\left(\frac{z - Z_l^\star}{b}\right) - K\left(\frac{z' - Z_l^\star}{b}\right) \right| \\ & \leq C n^{2\tau-1} b^{-1-d} |I_i| |z - z'|_\infty, \end{aligned}$$

where $|I_i|$ is the cardinality of I_i , which also entails that

$$|\mathbb{E}f_{i,jk}^\diamond(z) - \mathbb{E}f_{i,jk}^\diamond(z')| \leq \mathbb{E}|f_{i,jk}^\diamond(z) - f_{i,jk}^\diamond(z')| \leq Cn^{2\tau-1}b^{-1-d}|I_i||z - z'|_\infty.$$

Using above relations, we have

$$\sup_{z \in \mathcal{D}_u} |f_{i,jk}^\diamond(z) - \mathbb{E}f_{i,jk}^\diamond(z)| \leq |f_{i,jk}^\diamond(z_u) - \mathbb{E}f_{i,jk}^\diamond(z_u)| + Cn^{2\tau-1}b^{-1-d}|I_i|N^{-1}.$$

Thus,

$$\sup_{z \in \mathcal{D}} |f_{i,jk}^\diamond(z) - \mathbb{E}f_{i,jk}^\diamond(z)| \leq \max_{1 \leq u \leq N} |f_{i,jk}^\diamond(z_u) - \mathbb{E}f_{i,jk}^\diamond(z_u)| + Cn^{2\tau-1}b^{-1-d}|I_i|N^{-1}.$$

By taking $N = n^{9d}$ and $|I_i| = \lceil n/m \rceil$, we have

$$\sup_{z \in \mathcal{D}} |f_{i,jk}^\diamond(z) - \mathbb{E}f_{i,jk}^\diamond(z)| \leq \max_{1 \leq u \leq N} |f_{i,jk}^\diamond(z_u) - \mathbb{E}f_{i,jk}^\diamond(z_u)| + C \frac{n^{2\tau-9d}b^{-1-d}}{\log n}. \quad (\text{B.11})$$

Let

$$Y_{l,jk,z}^\star = X_{lj}^\star X_{lk}^\star K\left(\frac{z - Z_l^\star}{b}\right) - \mathbb{E}\left[X_{lj}^\star X_{lk}^\star K\left(\frac{z - Z_l^\star}{b}\right)\right], \quad l = 1, \dots, |I_i|.$$

Then, by the definition, we know $Y_{l,jk,z}^\star$ are i.i.d with $\mathbb{E}Y_{l,jk,z}^\star = 0$ and

$$|Y_{l,jk,z}^\star| \leq Cn^{2\tau}, \quad v = \sum_{l \in I_i} \mathbb{E}|Y_{l,jk,z}^\star|^2 \leq C|I_i|n^{2\tau}b^d.$$

By Bernstein's inequality and the new notation above, we then can bound the tail probability

for $f_{i,jk}^\diamond(z) - \mathbb{E}f_{i,jk}^\diamond(z) = (nb^d)^{-1} \sum_l Y_{l,jk,z}^\star$,

$$\mathbb{P}\left(\left|\frac{1}{nb^d} \sum_{l \in I_i} Y_{l,jk,z}^\star\right| > t, \max_i |X_i^\star|_\infty \leq n^\tau\right) \leq 2 \exp\left(-\frac{Ct^2 n^2 b^{2d}}{v + tb^d n^{2\tau+1}}\right). \quad (\text{B.12})$$

By taking $t = c_2 n^{\tau-1/2} b^{-d/2}$, where c_2 is a sufficient large constant such that

$$n^{9d} \exp \left(-\frac{Ct^2 n b^{2d} m}{n^{2\tau} b^d} \right) = n^{9d-C} = o(n^{-1}).$$

Consequently, for the chosen t , we have:

$$\begin{aligned} & \mathbb{P} \left(\max_{1 \leq u \leq N} \left| \frac{1}{n b^d} \sum_{l \in I_i} Y_{l,jk,z_u}^* \right| > t, \max_i |X_i^*|_\infty \leq n^\tau \right) \\ & \leq N \exp \left(-\frac{Ct^2 n^2 b^{2d}}{v} \right) + N \exp \left(-\frac{Ct^2 n^2 b^{2d}}{t b n^{2\tau+1}} \right) \\ & = o(n^{-1}). \end{aligned} \tag{B.13}$$

Combining (B.10), (B.11), (B.13) and $n^{2\tau-9d} b^{-1-d} / \log n = o(n^{\tau-1/2} b^{-d/2})$, we have

$$\mathbb{P} \left(\sup_z |f_n^*(z) - \mathbb{E} f_n^*(z)|_\infty > C n^{\tau-1/2} b^{-d/2} \log n, \max_i |X_i^*|_\infty \leq n^\tau \right) = o(n^{-1} \log n),$$

which together with (B.6), (B.7) and (B.9) proves the lemma. $\square$

B.2.2 Proof of Theorem 3.2.1

To state Lemma B.2.3, we remind that the first and the second derivatives of the coefficient function $\beta(z) : \mathbb{R}^d \rightarrow \mathbb{R}^p$ are a $p \times d$ matrix (order-2 tensor) and a $p \times d \times d$ order-3 tensor, respectively:

$$\begin{aligned} \nabla \beta(z) &= \left(\frac{\partial}{\partial z_k} \beta_j(z) \right)_{j=1, \dots, p; k=1, \dots, d} \in \mathbb{R}^{p \times d} \\ \nabla^2 \beta(z) &= \left(\frac{\partial^2}{\partial z_k \partial z_l} \beta_j(z) \right)_{j=1, \dots, p; k=1, \dots, d; l=1, \dots, d} \in \mathbb{R}^{p \times d \times d}. \end{aligned}$$

Recall the definition of k -mode product of an order- K tensor $\mathcal{X} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ with a vector $a \in \mathbb{R}^{d_k}$. Assume that

$$\mathcal{X} = (x_{i_1 \dots i_k})_{i_1=1, \dots, d_1; \dots; i_k=1, \dots, d_k}, \quad a = (a_{i_k})_{i_k=1, \dots, d_k}.$$

Then k -mode product of $\mathcal{X}$ with a (denoted as $\mathcal{X} \times_k a$) is an order- $(K-1)$ tensor of size $d_1 \times \cdots \times d_{k-1} \times d_{k+1} \times \cdots \times d_K$ and element-wisely,

$$(\mathcal{X} \times_k a)_{i_1 \dots i_{k-1} i_{k+1} \dots i_K} = \sum_{i_k=1}^{d_k} x_{i_1 \dots i_{k-1} i_k i_{k+1} \dots i_K} a_{i_k}.$$

Lemma B.2.3. *Let*

$$f_n^1(z) = \frac{1}{nb^d} \sum_{i=1}^n X_i X_i^\top \nabla \beta(z) \frac{z - Z_i}{b} K\left(\frac{z - Z_i}{b}\right),$$

$$f_n^2(z) = \frac{1}{nb^d} \sum_{i=1}^n X_i X_i^\top \nabla^2 \beta(z) \times_2 \frac{z - Z_i}{b} \times_3 \frac{z - Z_i}{b} K\left(\frac{z - Z_i}{b}\right).$$

Then under the same conditions of Lemma B.2.2, we have the same bound as (B.4):

$$\mathbb{P} \left(\sup_{z \in \mathcal{D}} |f_n^\alpha(z) - \mathbb{E} f_n^\alpha(z)|_\infty > C n^{1/q-1/2} b^{-d/2} \log n \right) = o(1). \quad (\text{B.14})$$

Proof. : Write $K_1^z(s) = \nabla \beta(z) s K(s)$ and $K_2^z(s) = \nabla^2 \beta(z) \times_2 s \times_3 s K(s)$ as new ‘kernels’ with dimension p . Because of Condition 3.2.1 and $s \in [-1, 1]^p$, the Lipschitz condition on each coordinate of K_1^z and K_2^z still holds. Then using the same proof of Lemma B.2.2, we have

$$f_n^\alpha(z) = \frac{1}{nb^d} \sum_{i=1}^n X_i X_i^\top K_\alpha \left(\frac{z - Z_i}{b} \right),$$

which readily leads to (B.14) for $\alpha = 1, 2$. □

Proof of Theorem 3.2.1: Throughout the proof, we use C to denote a generic constant, whose value may vary from line to line. Referring to the definition of $f_n(z)$ and $f_n^\alpha(z)$, to make the bandwidth b to be specified, we add b into subscript. Plugging in the explicit form (3.6) of $\hat{\beta}(z)$, we get the following decomposition

$$\begin{aligned} \frac{c^\top(\hat{\beta}_b(z) - \beta(z))}{\gamma(z)} &= c^\top f_{n,b}(z)^{-1} \frac{1}{nb^d} \sum_{i=1}^n X_i X_i^\top K\left(\frac{z - Z_i}{b}\right) (\beta(Z_i) - \beta(z)) / \gamma(z) \\ &\quad + c^\top f_{n,b}(z)^{-1} \frac{1}{nb^d} \sum_{i=1}^n X_i K\left(\frac{z - Z_i}{b}\right) \sigma(Z_i) \varepsilon_i / \gamma(z) \\ &= M_b(z) + R_b(z), \end{aligned} \tag{B.15}$$

where $M_b(z)$ and $R_b(z)$ represent the bias and the stochastic parts, respectively. To deal with the debiased estimator $\tilde{\beta}_b(z)$, by definition, we have

$$\begin{aligned} \frac{c^\top(\tilde{\beta}_b(z) - \beta(z))}{\gamma(z)} &= 2 \frac{c^\top(\hat{\beta}_b(z) - \beta(z))}{\gamma(z)} - \frac{c^\top(\hat{\beta}_{\sqrt{2}b}(z) - \beta(z))}{\gamma(z)} \\ &= 2M_b(z) - M_{\sqrt{2}b}(z) + 2R_b(z) - R_{\sqrt{2}b}(z), \end{aligned}$$

For the first bias correction part, apply Taylor's expansion on $\beta(z)$ to get

$$\begin{aligned} 2M_b(z) - M_{\sqrt{2}b}(z) &= \frac{c^\top}{\sigma(z)} \left[2f_{n,b}(z)^{-1} \left(f_{n,b}^1(z)b + \frac{1}{2}f_{n,b}^2(z)b^2 \right. \right. \\ &\quad \left. \left. + O(b^3) \right) - f_{n,\sqrt{2}b}(z)^{-1} \left(\sqrt{2}f_{n,\sqrt{2}b}^1(z)b + f_{n,\sqrt{2}b}^2(z)b^2 + O(b^3) \right) \right]. \end{aligned} \tag{B.16}$$

By Lemma B.2.2, Condition 3.2.3 and some elementary calculations, we have

$$\begin{aligned} f_{n,b}(z) &= \mathbb{E}f_{n,b}(z) + O_P(n^{1/q-1/2}b^{-d/2} \log n) \\ &= V(z) + O_P\left(b^2 + n^{1/q-1/2}b^{-d/2} \log n\right), \end{aligned} \tag{B.17}$$

where $V(\cdot)$ is defined in Condition 3.2.5. Similarly, by Lemma B.2.3, we have that

$$\begin{aligned} f_{n,b}^1(z) &= b \int \nabla V(z) \times_3 \tau \nabla \beta(z) \tau K(\tau) d\tau + O_P \left(n^{1/q-1/2} b^{-d/2} \log n + b^3 \right), \\ f_{n,b}^2(z) &= \int V(z) \nabla^2 \beta(z) \times_2 \tau \times_3 \tau K(\tau) d\tau + O_P \left(n^{1/q-1/2} b^{-d/2} \log n + b^2 \right). \end{aligned}$$

Plug the approximation of $f_{n,b}(z)$, $f_{n,b}^1(z)$ and $f_{n,b}^2(z)$ into (B.16) and by Condition 3.2.1 and Condition 3.2.5, we can bound $M(z)$ by

$$|2M_b(z) - M_{\sqrt{2}b}(z)| = O_P \left(b^3 + n^{1/q-1/2} b^{-d/2+1} \log n \right). \quad (\text{B.18})$$

For the second part $R(z) = 2R_b(z) - R_{\sqrt{2}b}(z) = \sum_{i=1}^n \tilde{w}_{i,b}^c(z) \varepsilon_i$, we first partition the domain $\mathcal{D}$ into N subdomains $\{\mathcal{D}_u, u = 1, \dots, N\}$ such that in each subdomain $\mathcal{D}_u, \forall z, z' \in \mathcal{D}_u, |z - z'|_\infty < C/N$ for some constant C independent of n and b . Let z_u be an arbitrary point of $\mathcal{D}_u$. By the Lipschitz condition on $w_{i,b}^c(z)$, we have that, with probability $1 - o(1)$,

$$|R(z) - R(z')| \leq C n^{1/q} b^{-1-d} |z - z'|_\infty.$$

This entails that

$$\sup_{z \in \mathcal{D}} |R(z)| \leq \max_{1 \leq u \leq N} |R(z_u)| + C n^{1/q} b^{-1-d} N^{-1}.$$

By taking $N = n^{2d}$, we have

$$\sup_{z \in \mathcal{D}} |R(z)| \leq \max_{1 \leq u \leq N} |R(z_u)| + C n^{1/q-2d} b^{-1-d}. \quad (\text{B.19})$$

Now we can do Gaussian approximation for $R(z)$. By Proposition 3.2.1, on a richer proba-

bility space there are $\phi_1, \dots, \phi_n \sim \text{i.i.d. } N(0, 1)$ such that

$$\begin{aligned} & \sup_{t>0} \left| \mathbb{P}^* \left(\max_{1 \leq u \leq N} |R(z_u)| > t \right) - \mathbb{P}^* \left(\max_{1 \leq u \leq N} \left| \sum_{i=1}^n \tilde{w}_{i,b}^c(z_u) \phi_i \right| > t \right) \right| \\ & \leq C_q (\log n)^{\frac{5q-6}{4q-2}} \left(\sum_{i=1}^n \max_{1 \leq u \leq N} \left| \frac{\tilde{w}_{i,b}^c(z_u)}{\sigma_u} \right|^q \right)^{\frac{1}{2q-1}}, \end{aligned} \quad (\text{B.20})$$

where $\sigma_u^2 = \mathbb{E}^* \left(\sum_{i=1}^n \tilde{w}_{i,b}^c(z_u) \varepsilon_i \right)^2$ and $\mathbb{E}^*$ is the expectation conditioning on X_i, Z_i . By the definition of $w_{i,b}^c(z_u)$ in (3.14), Lemma B.2.2 and Condition 3.2.5, with probability $1 - o(1)$, we have the following bound:

$$\begin{aligned} \sum_{i=1}^n w_{i,b}^c(z_u)^2 &= \sum_{i=1}^n \left[c^\top f_n(z_u)^{-1} \frac{1}{nb^d} X_i K \left(\frac{z_u - Z_i}{b} \right) \sigma(Z_i) / \gamma(z_u) \right]^2 \\ &= \frac{1}{n^2 b^{2d}} c^\top \left[V(z_u) + O(n^{1/q-1/2} b^{-d/2}) \right]^{-1} \sum_{i=1}^n X_i X_i^\top K \left(\frac{z_u - Z_i}{b} \right)^2 \\ &\quad \times \left[V(z_u) + O(n^{1/q-1/2} b^{-d/2}) \right]^{-1} c \left(\frac{\sigma(Z_i)}{\gamma(z_u)} \right)^2 \\ &\geq \frac{C}{nb^d}. \end{aligned}$$

Additionally, we have exactly the same lower bound for $\sum_{i=1}^n w_{i,\sqrt{2}b}^c(z_u)^2$, which derives the bound for σ_u ,

$$\sigma_u^2 = \sum_{i=1}^n \left(2w_{i,b}^c(z_u) - w_{i,\sqrt{2}b}^c(z_u) \right)^2 \geq \frac{C}{nb^d}.$$

For the same reason, we can bound $\max_u |w_i(z_u)|$ by

$$\begin{aligned} \max_{1 \leq u \leq N} |w_{i,b}^c(z_u)| &= \max_u \frac{1}{nb^d} c^\top \left[V(z_u) + O(n^{1/q-1/2} b^{-d/2}) \right]^{-1} X_i K \left(\frac{z_u - Z_i}{b} \right) \frac{\sigma(Z_i)}{\gamma(z_u)} \\ &\leq \frac{C}{nb^d} |X_i|_\infty, \end{aligned}$$

with probability $1 - o(1)$, and we have the same upper bound for $\max_u |w_{i,\sqrt{2}b}^c(z_u)|$. So, by

the tightness,

$$\sum_{i=1}^n \max_{1 \leq u \leq N} \left| \frac{\tilde{w}_{i,b}^c(z_u)}{\sigma_u} \right|^q \leq \sum_{i=1}^n C \frac{|X_i|_\infty^q}{n^{q/2} b^{qd/2}} = O_P \left(\frac{n}{n^{q/2} b^{qd/2}} \right),$$

which derives a bound for (B.19)

$$\begin{aligned} & \sup_{t>0} \left| \mathbb{P}^* \left(\max_{1 \leq u \leq N} |R(z_u)| > t \right) - \mathbb{P}^* \left(\max_{1 \leq u \leq N} \left| \sum_{i=1}^n \tilde{w}_{i,b}^c(z_u) \phi_i \right| > t \right) \right| \\ & \leq C (\log n)^{\frac{5q-6}{4q-2}} n^{\frac{2-q}{4q-2}} b^{\frac{-qd}{4q-2}}. \end{aligned} \quad (\text{B.21})$$

Meanwhile, with the exactly the same reason and grid, we have

$$\sup_{z \in \mathcal{D}} \left| \sum_{i=1}^n \tilde{w}_{i,b}^c(z) \phi_i \right| \leq \max_{1 \leq u \leq N} \left| \sum_{i=1}^n \tilde{w}_{i,b}^c(z_u) \phi_i \right| + C n^{1/q-2d} b^{-1-d}.$$

Comparing the above probability with the one shown in the bound (B.4),

$$1 - n^{1-\tau q} \mathbb{E} \left[|X_i|_\infty^q \mathbf{1} \{ |X_i|_\infty \geq n^\tau \} \right] = 1 - o(1),$$

which together with (B.15), (B.18), (B.19), (B.20) proves the theorem. $\square$

B.2.3 Proof of Theorem 3.2.2

Proof of Theorem 3.2.2: Recall

$$\hat{\sigma}^2(z) = \left(\sum_{t=1}^n K_h(z - Z_t) \right)^{-1} \sum_{t=1}^n K_h(z - Z_t) \left(Y_t - \tilde{\beta}(Z_t)^\top X_t \right)^2.$$

Let $f_{n1}(z) = \frac{1}{nh^d} \sum_t K_h(z - Z_t)$. Plug in $Y_t = \beta(Z_t)^\top X_t + \sigma(Z_t)\varepsilon_t$ and we have

$$\begin{aligned}
\hat{\sigma}^2(z) &= \frac{1}{nh^d f_{n1}(z)} \sum_{t=1}^n K_h(z - Z_t) [\sigma(Z_t)\varepsilon_t]^2 \\
&\quad + \frac{2}{nh^d f_{n1}(z)} \sum_{t=1}^n K_h(z - Z_t) [\beta(Z_t) - \tilde{\beta}(Z_t)]^\top X_t \sigma(Z_t) \varepsilon_t \\
&\quad + \frac{1}{nh^d f_{n1}(z)} \sum_{t=1}^n K_h(z - Z_t) [(\beta(Z_t) - \tilde{\beta}(Z_t))^\top X_t]^2 \\
&= \sigma_{n1}^2(z) + c_{n2}(z) + \sigma_{n3}^2(z).
\end{aligned} \tag{B.22}$$

For $\sigma_{n1}^2(z)$,

$$\begin{aligned}
(\sigma_{n1}^2(z) - \sigma^2(z))f_{n1}(z) &= \frac{1}{nh^d} \sum_{t=1}^n K_h(z - Z_t) \sigma(Z_t)^2 (\varepsilon_t^2 - 1) \\
&\quad + \frac{1}{nh^d} \sum_{t=1}^n K_h(z - Z_t) [\sigma^2(Z_t) - \sigma^2(z)] \\
&= M_{n1}(z) + R_{n2}(z).
\end{aligned}$$

Let $\tilde{\varepsilon}_t = \varepsilon_t^2 - 1 \in \mathcal{L}^{q/2}$. Then, $\max_t |\tilde{\varepsilon}_t| = O_P(n^{2/q})$. With typically the same idea as in (B.10), we can do the partition and gridding so that

$$\mathbb{P} \left(\sup_z |M_{n1}(z)| > t \right) \leq \sum_{i=1}^m \mathbb{P} \left(\sup_z \left| \frac{1}{nh^d} \sum_{l \in I_i} K_h(z - Z_l) \tilde{\varepsilon}_l \right| > \frac{t}{m} \right).$$

Because of independence between Z_t and ε_t , we have

$$v = \sum_{i \in I_i} \mathbb{E} K_h^2(z - Z_l) \tilde{\varepsilon}_l^2 = O(h^d).$$

Conditioning on $\max_t |\tilde{\varepsilon}_t| \leq n^{2/q}$, by Bernstein's inequality, we have

$$\mathbb{P} \left(\sup_z |M_{n1}(z)| > Cn^{2/q-1}h^{-d} \log n \right) \leq n^{-1}.$$

Therefore

$$\sup_{0 \leq z \leq 1} |M_{n1}(z)| = O_P(n^{2/q-1}h^{-d} \log n). \quad (\text{B.23})$$

Note that $\sup_z |K_h(z - Z_t)[\sigma^2(Z_t) - \sigma^2(z)]| \leq Ch$. Then

$$\sup_z |R_{n2}(z)| = O_P(h). \quad (\text{B.24})$$

For the interaction term $c_{n2}(z)$,

$$\begin{aligned} & \sum_{t=1}^n K_h(z - Z_t) [\beta(Z_t) - \hat{\beta}(Z_t)]^\top X_t \sigma(Z_t) \varepsilon_t \\ &= \frac{1}{nb^d} \sum_{t=1}^n \sum_{i=1}^n K_h(z - Z_t) K_b(Z_t - Z_i) \sigma(Z_t) \sigma(Z_i) \varepsilon_t \varepsilon_i [\beta(Z_i) - \beta(Z_t)]^\top X_i X_i^\top f_{n,b}(Z_t)^{-1} X_t \\ & \quad + \frac{1}{nb^d} \sum_{t=1}^n \sum_{i=1}^n K_h(z - Z_t) K_b(Z_t - Z_i) \sigma(Z_t) \sigma(Z_i) \varepsilon_t \varepsilon_i X_i^\top f_{n,b}^{-1}(Z_t) X_t. \end{aligned}$$

Notice that $\text{supp}(K) = [-1, 1]$ and $\beta \in \mathcal{C}^3(\mathcal{D})$. By Lemma 1, we have

$$\mathbb{E}[X_i^\top f_{n,b}^{-1}(Z_t) X_t] \leq C \|X_i\| \|X_t\| = O(1), \quad \mathbb{E} \left\{ [\beta(Z_i) - \beta(Z_t)]^\top X_i \right\} = O(b).$$

The remaining part is similar as that in Zhao and Wu [2008c], and we also have

$$\sup_{0 \leq z \leq 1} |c_{n2}(z)| = O_p \left(\frac{1}{nb^{5d/2}} \right). \quad (\text{B.25})$$

By the proof the Theorem 3.2.1, we have

$$\sup_{0 \leq z \leq 1} |\sigma_{n3}^2(z)| = O_P \left(\frac{\log n}{nb^d} + b^6 + n^{2/q-1} b^{-d+2} (\log n)^2 \right). \quad (\text{B.26})$$

Combining (B.23)-(B.26) and after simplification, we can derive the bound for $|\hat{\sigma}^2(z) - \sigma^2(z)|$. $\square$

B.2.4 Proof of Theorem 3.2.3

Proof of Theorem 3.2.3: Following the proof for Theorem 3.2.1, we can derive the approximation based on grid for both $Y_i = \beta(Z_i)^\top X_i + \sigma(Z_i)\varepsilon_i$ and $Y_i^\star = \hat{\sigma}(Z_i)\phi_i$ and choose the same bandwidth b and $1/n^{2d}$ -net $\{z_1, \dots, z_N\}$. Then the problem turn to be showing the bound of

$$\sup_{u>0} \left| \mathbb{P}^* \left(\left| \sum_{i=1}^n W_i \phi_{1,i} \right|_\infty > u \right) - \mathbb{P}^* \left(\left| \sum_{i=1}^n W_i^\star \phi_{2,i} \right|_\infty > u \right) \right|, \quad (\text{B.27})$$

where $\phi_{1,i}$ and $\phi_{2,i}$ are i.i.d. $N(0, 1)$ for $i = 1, \dots, n$. In contrast to (3.14),

$$w_{i,b}^{c,\star}(z) = c^\top \left(\sum_i X_i X_i^\top K \left(\frac{z - Z_i}{b} \right) \right)^{-1} X_i K \left(\frac{z - Z_i}{b} \right) \hat{\sigma}(Z_i) / \gamma(z).$$

Hence W_i and $W_i^\star$ are defined by

$$W_i = \left(\tilde{w}_{i,b}^c(z_1), \dots, \tilde{w}_{i,b}^c(z_N) \right)^\top, \quad W_i^\star = \left(\tilde{w}_{i,b}^{c,\star}(z_1), \dots, \tilde{w}_{i,b}^{c,\star}(z_N) \right)^\top$$

$$\tilde{w}_{i,b}^c(z_j) = 2w_{i,b}^c(z_j) - w_{i,\sqrt{2}b}^c(z_j), \quad \tilde{w}_{i,b}^{c,\star}(z_j) = 2w_{i,b}^{c,\star}(z_j) - w_{i,\sqrt{2}b}^{c,\star}(z_j),$$

for $i = 1, \dots, n$ and $j = 1, \dots, N$. Without loss of recognition, we denote $\tilde{w}_i(z_j) = \tilde{w}_{i,b}^c(z_j)$, $\tilde{w}_i^\star(z_j) = \tilde{w}_{i,b}^{c,\star}(z_j)$, $w_{i,b}(z_j) = w_{i,b}^c(z_j)$ and $w_{i,b}^\star(z_j) = w_{i,b}^{c,\star}(z_j)$. Let $Z_1 = \sum_{i=1}^n \sqrt{nb^d} W_i \phi_{1,i}$, $Z_2 = \sqrt{nb^d} \sum_{i=1}^n W_i^\star \phi_{2,i}$, $\Sigma^1 = \text{Cov}(Z_1)$ and $\Sigma^2 = \text{Cov}(Z_2)$. We can show the explicit form

of each element of covariance matrices

$$\Sigma_{jk}^1 = nb^d \sum_{i=1}^n \tilde{w}_i(z_j) \tilde{w}_i(z_k), \quad \Sigma_{jk}^2 = nb^d \sum_{i=1}^n \tilde{w}_i^*(z_j) \tilde{w}_i^*(z_k).$$

By the definition of $w_{i,b}^c(z)$, we have

$$\begin{aligned} \Sigma_{jj}^1 &= \sum_{i=1}^n \left[\sqrt{nb^d} \tilde{w}_i(z_j) \right]^2 \\ &= nb^d \sum_{i=1}^n c^\top \left(2 \left(\sum_{s=1}^n X_s X_s^\top K \left(\frac{z_j - Z_s}{b} \right) \right)^{-1} X_i K \left(\frac{z_j - Z_i}{b} \right) \right. \\ &\quad \left. - \left(\sum_{s=1}^n X_s X_s^\top K \left(\frac{z_j - Z_s}{\sqrt{2}b} \right) \right)^{-1} X_i K \left(\frac{z_j - Z_i}{\sqrt{2}b} \right) \right) \left(\frac{\sigma(Z_i)}{\gamma(z_j)} \right)^2 \\ &\quad \left(2 \left(\sum_{s=1}^n X_s X_s^\top K \left(\frac{z_j - Z_s}{b} \right) \right)^{-1} X_i K \left(\frac{z_j - Z_i}{b} \right) \right. \\ &\quad \left. - \left(\sum_{s=1}^n X_s X_s^\top K \left(\frac{z_j - Z_s}{\sqrt{2}b} \right) \right)^{-1} X_i K \left(\frac{z_j - Z_i}{\sqrt{2}b} \right) \right)^\top c \\ &\geq nb^d \left(\frac{M}{m'} \right)^2 \sum_{i=1}^n c^\top \left(2f_{n,b}^{-1}(z_j) \frac{1}{nb^d} X_i K \left(\frac{z_j - Z_i}{b} \right) - \right. \\ &\quad \left. f_{n,\sqrt{2}b}^{-1}(z_j) \frac{1}{\sqrt{2}nb^d} X_i K \left(\frac{z_j - Z_i}{\sqrt{2}b} \right) \right) \left(2f_{n,b}^{-1}(z_j) \frac{1}{nb^d} X_i K \left(\frac{z_j - Z_i}{b} \right) \right. \\ &\quad \left. - f_{n,\sqrt{2}b}^{-1}(z_j) \frac{1}{\sqrt{2}nb^d} X_i K \left(\frac{z_j - Z_i}{\sqrt{2}b} \right) \right)^\top c \\ &= \left(\frac{M}{m'} \right)^2 c^\top \sum_{i=1}^n \left(4f_{n,b}^{-2}(z_j) \frac{1}{(nb^d)^2} X_i X_i^\top K \left(\frac{z_j - Z_i}{b} \right)^2 \right. \\ &\quad \left. - 4f_{n,b}^{-1}(z_j) f_{n,\sqrt{2}b}^{-1}(z_j) \frac{1}{\sqrt{2}(nb^d)^2} X_i X_i^\top K \left(\frac{z_j - Z_i}{b} \right) K \left(\frac{z_j - Z_i}{\sqrt{2}b} \right) \right. \\ &\quad \left. + f_{n,\sqrt{2}b}^{-2}(z_j) \frac{1}{2(nb^d)^2} X_i X_i^\top K \left(\frac{z_j - Z_i}{\sqrt{2}b} \right)^2 \right) c. \end{aligned}$$

Regard $K((z - Z_i)/b)^2$, $K((z - Z_i)/b)K((z - Z_i)/\sqrt{2}b)$ and $K((z - Z_i)/\sqrt{2}b)^2$ as new kernels and apply Lemma B.2.2 to plug (B.17) into the above bound, then we have

$$\Sigma_{jj}^1 \geq \left(\frac{M}{m'}\right)^2 c^\top \frac{9 - 4\sqrt{2}}{2} \left(V(z_j) + O_P(n^{1/q-1/2}b^{-d/2} \log n + b^2)\right)^{-1} c. \quad (\text{B.28})$$

It's easy to find that the supremum of the maximum eigenvalue of $V(z)$, $z \in \mathcal{D}$ is bounded by $\mathbb{E}X_i^\top X_i \sup_z f(z)$. So Σ_{jj}^1 has a constant uniformly lower bound for all $j = 1, \dots, N$.

Now consider

$$\begin{aligned} \Delta &= \max_{1 \leq j, k \leq N} |\Sigma_{jk}^1 - \Sigma_{jk}^2| \\ &= \max_{1 \leq j, k \leq N} nb^d \sum_{i=1}^n |\tilde{w}_i(z_j) \tilde{w}_i(z_k) - \tilde{w}_i^*(z_j) \tilde{w}_i^*(z_k)| \\ &= \max_{j, k} nb^d \sum_{i=1}^n c^\top \left(2 \left(\sum_s X_s X_s^\top K \left(\frac{z_j - Z_s}{b} \right) \right)^{-1} X_i K \left(\frac{z_j - Z_i}{b} \right) \right. \\ &\quad \left. - \left(\sum_s X_s X_s^\top K \left(\frac{z_j - Z_s}{\sqrt{2}b} \right) \right)^{-1} X_i K \left(\frac{z_j - Z_i}{\sqrt{2}b} \right) \right. \\ &\quad \left. \left(2 \left(\sum_s X_s X_s^\top K \left(\frac{z_k - Z_s}{b} \right) \right)^{-1} X_i K \left(\frac{z_k - Z_i}{b} \right) \right. \right. \\ &\quad \left. \left. - \left(\sum_s X_s X_s^\top K \left(\frac{z_k - Z_s}{\sqrt{2}b} \right) \right)^{-1} X_i K \left(\frac{z_k - Z_i}{\sqrt{2}b} \right) \right) \right)^\top c \\ &\quad \left| \frac{\sigma^2(Z_i)}{\gamma(z_j)\gamma(z_k)} - \frac{\hat{\sigma}^2(Z_i)}{\gamma(z_j)\gamma(z_k)} \right|. \end{aligned}$$

By Theorem 3.2.2, we can bound the covariance part with

$$\left| \frac{\sigma^2(Z_i)}{\gamma(z_j)\gamma(z_k)} - \frac{\hat{\sigma}^2(Z_i)}{\gamma(z_j)\gamma(z_k)} \right| \leq O_P(\tau_n). \quad (\text{B.29})$$

With similar analysis, the coefficient part before the covariance part has a constant upper bound with probability $1 - o(1)$, which derives $\Delta \leq O_P(\tau_n)$. By Gaussian-to-Gaussian comparison proposition in Chernozhuokov et al. [2022], we can bound (B.27) by

$$\sup_{u>0} \left| \mathbb{P}^* \left(\left| \sum_{i=1}^n W_i \phi_{1,i} \right|_{\infty} > u \right) - \mathbb{P}^* \left(\left| \sum_{i=1}^n W_i^* \phi_{2,i} \right|_{\infty} > u \right) \right| \leq O(\tau_n),$$

which leads to the final consequence. □

APPENDIX C

APPENDIX: SUPPLEMENTARY MATERIALS FOR CHAPTER

4

C.1 Additional Figures

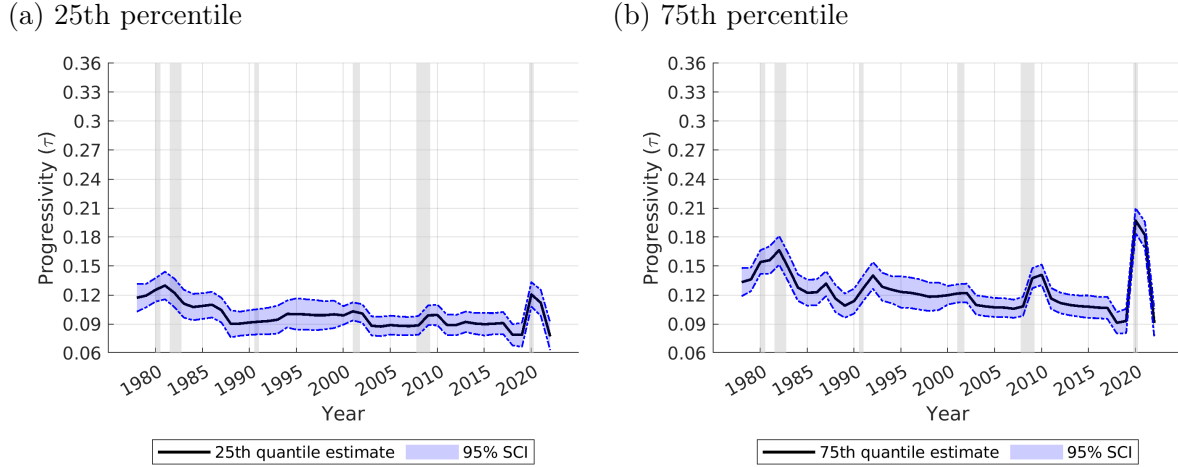


Figure C.1: Progressivity Estimates and Simultaneous Confidence Intervals

Notes: Panels (a) and (b) show the U.S. tax progressivity estimates at the 25th and 75th income percentiles (black solid curve) and 95% simultaneous confidence bands obtained using the simultaneous inference method (blue area). The grey-shaded regions indicate NBER recessions.

C.2 Technical Proofs

To prove Theorem and Corollary, we firstly present a lemma on the concentration of $\hat{\beta}$.

Lemma C.2.1. *Under the assumptions of the Theorem, we have,*

$$\mathbb{P} \left(\sup_{|\beta - \beta_0| \leq \delta} |K(\beta) - K(\beta_0)| \geq t \right) \leq c_1 (\delta n)^{3p} \exp \left(-\frac{c_2 t^2}{n\delta + n^{1/3}t} \right) + c_3 n^{-1/3}$$

where c_1, c_2 are constants not depending on n .

Proof. Throughout the proof, we use C, c_1 and c_2 to denote generic constants not depending on n , which may vary from line to line. Let $\Delta = \beta - \beta_0$. By the definition of $K(\beta)$, we have

$$\begin{aligned} K(\beta) - K(\beta_0) &= \sum_{i=1}^n (\mathbb{1}_{y_i \leq x_i^\top \beta} - \mathbb{E} \mathbb{1}_{y_i \leq x_i^\top \beta}) x_i - \sum_{i=1}^n (\mathbb{1}_{y_i \leq x_i^\top \beta_0} - \mathbb{E} \mathbb{1}_{y_i \leq x_i^\top \beta_0}) x_i \\ &:= \sum_{i=1}^n Z_i, \end{aligned}$$

where $Z_i = (\mathbb{1}_{y_i \leq x_i^\top \beta} - \mathbb{1}_{y_i \leq x_i^\top \beta_0} - \mathbb{E} \mathbb{1}_{y_i \leq x_i^\top \beta} + \mathbb{E} \mathbb{1}_{y_i \leq x_i^\top \beta_0}) x_i$. Let $A_n = \{\max_{i \leq n} |X_i| \leq n^{1/3}\}$. By Markov inequality, we have

$$\mathbb{P}(\max_i |x_i| > n^{1/3}) \leq \frac{n}{n^{4/3}} \mathbb{E}[|x_i|^4 \mathbb{1}_{|x_i| > n^{1/3}}] = o(n^{-1/3}),$$

which indicates that with probability $1 - o(n^{-1/3})$,

$$\max_i |x_i| \leq n^{1/3} \text{ and } \max_i |Z_i| \leq 2n^{1/3}. \quad (\text{C.1})$$

Since (x_i, y_i) are i.i.d., Z_i are also i.i.d. and

$$\begin{aligned} \mathbb{E} Z_{ij}^2 &= \mathbb{E}[\mathbb{E}[(\mathbb{1}_{e_i \leq x_i^\top \Delta} - \mathbb{1}_{e_i \leq 0} - \mathbb{E} \mathbb{1}_{e_i \leq x_i^\top \Delta} + \mathbb{E} \mathbb{1}_{e_i \leq 0})^2 x_{ij}^2 | x_i]] \\ &= \mathbb{E}[\mathbb{E}[\mathbb{1}_{(e_i - x_i^\top \Delta) e_i \leq 0}](1 - \mathbb{E}[\mathbb{1}_{(e_i - x_i^\top \Delta) e_i \leq 0}]) | x_i] \mathbb{E} x_{ij}^2] \\ &\leq \mathbb{E}[\mathbb{E}[\mathbb{1}_{(e_i - x_i^\top \Delta) e_i \leq 0} | x_i] \mathbb{E} x_{ij}^2] \\ &\leq \sup_x f_{e_i | x_i}(x) |\Delta| \|x_i\|^3. \end{aligned} \quad (\text{C.2})$$

Therefore, we have

$$V = \sum_{i=1}^n \|Z_i\|^2 = \sum_{i=1}^n \sum_{j=1}^p \mathbb{E} Z_{ij}^2 \leq Cn|\Delta|. \quad (\text{C.3})$$

By vector Bernstein's inequality (Tropp et al. [2015]), we have

$$\begin{aligned} & \mathbb{P}(|K(\beta) - K(\beta_0)| > t) \\ & \leq \mathbb{P}(|K(\beta) - K(\beta_0)| > t, A_n) + \mathbb{P}(A_n^c) \\ & \leq (p+1) \exp\left(-\frac{c_1 t^2}{n|\Delta| + n^{1/3}t}\right) + c_2 n^{-1/3}. \end{aligned} \quad (\text{C.4})$$

For any $\delta > 0$, we can partition the compact domain $\mathcal{B}_{\delta, \beta_0}^p = \{\beta \in \mathbb{R}^p : |\beta - \beta_0| \leq \delta\}$ into $N = (2\delta pn)^{3p}$ subdomain $\mathcal{D}_i$, $i = 1, \dots, N$ such that

$$\mathcal{B}_{\delta, \beta_0}^p = \bigcup_{i=1}^N \mathcal{D}_i, \text{ and } |\beta - \beta'| \leq n^{-3} \text{ for any } \beta, \beta' \in \mathcal{D}_i.$$

Consider

$$K(\beta) - K(\beta') = \sum_{i=1}^n (\mathbb{1}_{y_i \leq x_i^\top \beta} - \mathbb{1}_{y_i \leq x_i^\top \beta'} - \mathbb{E} \mathbb{1}_{y_i \leq x_i^\top \beta} - \mathbb{E} \mathbb{1}_{y_i \leq x_i^\top \beta'}) x_i.$$

For the similar reason as (C.2), we have that if $|\beta - \beta'| \leq n^{-3}$, then, under A_n , we have

$$\begin{aligned} \mathbb{P}(\mathbb{1}_{y_i \leq x_i^\top \beta} - \mathbb{1}_{y_i \leq x_i^\top \beta'} \neq 0) &= \mathbb{P}((y_i - x_i^\top \beta)(y_i - x_i^\top \beta') < 0) \\ &= \mathbb{E} \left| \int_{x_i^\top (\beta - \beta_0)}^{x_i^\top (\beta' - \beta_0)} f_{e_i|x_i}(u) du \right| \\ &\leq M \|x_i\| |\beta - \beta'|, \end{aligned}$$

which indicates

$$|K(\beta) - K(\beta')| \leq Cn^{-5/3}.$$

Choose one arbitrary element β_i from each $\mathcal{D}_i$, $i = 1, \dots, N$. Let β be an arbitrary element in $\mathcal{B}_{\delta, \beta_0}^p$. Then there exists $\mathcal{D}_i$ such that $\beta \in \mathcal{D}_i$ so that $|\beta - \beta_i| \leq n^{-3}$. With this intention, under A_n , we have

$$\sup_{\beta \in \mathcal{B}_{\delta, \beta_0}^p} |K(\beta) - K(\beta_0)| \leq \max_{i=1, \dots, N} |K(\beta_i) - K(\beta_0)| + Cn^{-5/3}.$$

By (C.4), we have,

$$\begin{aligned} \mathbb{P}\left(\sup_{|\beta - \beta_0| \leq \delta} |K(\beta) - K(\beta_0)| \geq t\right) &\leq \mathbb{P}\left(\max_{i=1, \dots, N} |K(\beta_i) - K(\beta_0)| + Cn^{-5/3} \geq t, A_n\right) + \mathbb{P}(A_n^c) \\ &\leq N(p+1) \exp\left(-\frac{Ct^2}{n\delta + n^{1/3}t}\right) + c_3 n^{-1/3} \\ &\leq c_1(\delta n)^{3p} \exp\left(-\frac{c_2 t^2}{n\delta + n^{1/3}t}\right) + c_3 n^{-1/3}, \end{aligned}$$

completing the proof. □

Lemma C.2.2. *Under the assumptions of the Theorem, we have*

$$\mathbb{P}(|\hat{\beta} - \beta_0| > n^{-1/3}) \leq Cn^{-1/3},$$

where C is a constance not depending on n .

Proof. Let

$$M_n(\beta) = \sum_{i=1}^n \rho_\alpha(y_i - x_i^\top \beta), \quad M(\beta) = \mathbb{E}[M_n(\beta)].$$

Then, by the definition of $\hat{\beta}$, we have $\hat{\beta} = \arg \min_{\beta} M_n(\beta)$. Remind that $A_n = \{\max_{i \leq n} |X_i| \leq n^{1/3}\}$. Then,

$$\mathbb{P}(|\hat{\beta} - \beta_0| \geq n^{1/3}) \leq \mathbb{P}\left(\inf_{|\beta - \beta_0| \geq n^{1/3}} M_n(\beta) - M_n(\beta_0) < 0, A_n\right) + \mathbb{P}(A_n^c). \quad (\text{C.5})$$

For the same reason as shown in the proof of Lemma C.2.1, we can bound $\mathbb{P}(A_n^c)$ by

$$\mathbb{P}(A_n^c) \leq Cn^{-1/3}.$$

Meanwhile, by the union bound and convexity of $M_n(\cdot)$, we have

$$\begin{aligned} \mathbb{P}\left(\inf_{|\beta-\beta_0|\geq n^{1/3}} M_n(\beta) - M_n(\beta_0) < 0, A_n\right) &\leq \mathbb{P}\left(\inf_{|\beta-\beta_0|=n^{1/3}} M_n(\beta) - M_n(\beta_0) < 0, A_n\right) \\ &\leq \mathbb{P}\left(\inf_{|\beta-\beta_0|=n^{1/3}} M(\beta) - M(\beta_0) < \varepsilon, A_n\right) \\ &\quad + \mathbb{P}\left(\sup_{|\beta-\beta_0|=n^{1/3}} |[M_n(\beta) - M(\beta)] - [M_n(\beta_0) - M(\beta_0)]| > \varepsilon, A_n\right) \end{aligned} \tag{C.6}$$

Notice that

$$M(\beta) - M(\beta_0) = \mathbb{E}[M_n(\beta) - M_n(\beta_0)]$$

is deterministic and a function of β and $\nabla_\beta M(\beta) = g(\beta) \in \mathcal{C}^2$. Then by Taylor's expansion, we have that, as $|\beta - \beta_0| \rightarrow 0$,

$$M(\beta) - M(\beta_0) = (\beta - \beta_0)^\top g(\beta_0) + \frac{1}{2}(\beta - \beta_0)^\top \nabla^2 M(\beta_0)(\beta - \beta_0) + o(n|\beta - \beta_0|^3),$$

where $g(\beta_0) = 0$ and $\nabla^2 M(\beta_0)$ is positive definite. Therefore,

$$\begin{aligned} \inf_{|\beta-\beta_0|=n^{1/3}} M(\beta) - M(\beta_0) &= \inf_{|\beta-\beta_0|=n^{1/3}} \frac{1}{2}(\beta - \beta_0)^\top \nabla^2 M(\beta_0)(\beta - \beta_0) + o(1) \\ &= c_1 n^{1/3} + o(1), \end{aligned}$$

where c_1 is a positive constant. Choose $\varepsilon = n^{1/3-v}$ for any $v > 0$, then, as $n > N$ for some

constant N , we have

$$\mathbb{P} \left(\inf_{|\beta - \beta_0| = n^{1/3}} M(\beta) - M(\beta_0) < n^{1/3-v}, A_n \right) = 0. \quad (\text{C.7})$$

Let $J(\beta) = M_n(\beta) - M(\beta)$. By the definition of $M_n(\beta)$, we have

$$J(\beta) - J(\beta_0) = \sum_{i=1}^n [\rho_\alpha(y_i - x_i^\top \beta) - \rho_\alpha(y_i - x_i^\top \beta_0) - \mathbb{E}[\rho_\alpha(y_i - x_i^\top \beta)] + \mathbb{E}[\rho_\alpha(y_i - x_i^\top \beta_0)]].$$

Let $Z_i = [\rho_\alpha(y_i - x_i^\top \beta) - \rho_\alpha(y_i - x_i^\top \beta_0) - \mathbb{E}[\rho_\alpha(y_i - x_i^\top \beta)] + \mathbb{E}[\rho_\alpha(y_i - x_i^\top \beta_0)]]$ and $\delta = |\beta - \beta_0|$. Notice that $|\rho_\alpha(x) - \rho_\alpha(y)| \leq |x - y|$ for any $x, y \in \mathbb{R}$ and $\alpha \in (0, 1)$. Under the condition $\max_{i \leq n} |x_i| \leq n^{1/3}$, we have

$$\max_{i \leq n} |\rho_\alpha(y_i - x_i^\top \beta) - \rho_\alpha(y_i - x_i^\top \beta_0)| \leq |x_i^\top (\beta - \beta_0)| \leq n^{1/3} \delta.$$

Therefore, under condition A_n ,

$$\max_{i \leq n} |Z_i| \leq 2n^{1/3} \delta.$$

Meanwhile,

$$\begin{aligned} \mathbb{E} Z_i^2 &= \mathbb{E} [\rho_\alpha(y_i - x_i^\top \beta) - \rho_\alpha(y_i - x_i^\top \beta_0)]^2 + 3[\mathbb{E} [\rho_\alpha(y_i - x_i^\top \beta) - \rho_\alpha(y_i - x_i^\top \beta_0)]]^2 \\ &\leq C \delta^2. \end{aligned}$$

Then by Bernstein's inequality, we have

$$\mathbb{P}(|J(\beta) - J(\beta_0)| \geq \varepsilon, A_n) \leq C \exp \left(\frac{c_1 \varepsilon^2}{n \delta^2 + n^{1/3} \delta \varepsilon} \right).$$

For $\delta = n^{-1/3}$, we can partition the sphere $\mathcal{S}_{\delta, \beta_0}^p = \{\beta \in \mathbb{R}^p : |\beta - \beta_0| = \delta\}$ into $N = (2\delta p n)^p$

subdomain $\mathcal{D}_i$, $i = 1, \dots, N$ such that

$$\mathcal{S}_{\delta, \beta_0}^p = \bigcup_{i=1}^N \mathcal{D}_i, \text{ and } |\beta - \beta'| \leq n^{-1} \text{ for any } \beta, \beta' \in \mathcal{D}_i.$$

Consider

$$J(\beta) - J(\beta') = \sum_{i=1}^n [\rho_\alpha(y_i - x_i^\top \beta) - \rho_\alpha(y_i - x_i^\top \beta') - \mathbb{E}[\rho_\alpha(y_i - x_i^\top \beta)] + \mathbb{E}[\rho_\alpha(y_i - x_i^\top \beta')]].$$

For the similar reason as (C.2), we have that if $|\beta - \beta'| \leq n^{-1}$, then, under condition A_n ,

$$|J(\beta) - J(\beta')| \leq Cn^{-2/3}.$$

Choose one arbitrary element β_i from each $\mathcal{D}_i$, $i = 1, \dots, N$. Let β be an arbitrary element in $\mathcal{S}_{\delta, \beta_0}^p$. Then there exists $\mathcal{D}_i$ such that $\beta \in \mathcal{D}_i$ so that $|\beta - \beta_i| \leq n^{-1}$. With this intention, we have

$$\sup_{\beta \in \mathcal{S}_{\delta, \beta_0}^p} |J(\beta) - J(\beta_0)| \leq \max_{i=1, \dots, N} |J(\beta_i) - J(\beta_0)| + Cn^{-2/3}.$$

By (C.4), we have

$$\begin{aligned} \mathbb{P}(\sup_{|\beta - \beta_0| = \delta} |J(\beta) - J(\beta_0)| \geq \varepsilon, A_n) &\leq \mathbb{P}(\max_{i=1, \dots, N} |J(\beta_i) - J(\beta_0)| + Cn^{-2/3} \geq \varepsilon, A_n) \\ &\leq CN(p+1) \exp\left(-\frac{c_1 \varepsilon^2}{n\delta^2 + n^{1/3}\delta\varepsilon}\right) \\ &\leq C(\delta n)^p \exp\left(-\frac{c_1 \varepsilon^2}{n\delta^2 + n^{1/3}\delta\varepsilon}\right). \end{aligned}$$

Plug in $\delta = n^{-1/3}$ and $\varepsilon = n^{1/3-v}$ with $0 < v < 1/6$, then

$$\mathbb{P}(\sup_{|\beta - \beta_0| = n^{-1/3}} |J(\beta) - J(\beta_0)| \geq n^{1/3-v}, A_n) \leq Cn^{2p/3} \exp(-n^{1/3-2v}),$$

which together with (C.5), (C.6) and (C.7) completes the proof. $\square$

Proof of Theorem 4.2.1: By the union bound and triangular inequality, we have, for any convex set A ,

$$\begin{aligned} |\mathbb{P}(n^{-1/2}(g_n(\hat{\beta}) - g(\hat{\beta})) \in A) - \mathbb{P}(N(0, \Gamma) \in A)| &\leq \mathbb{P}\left(\frac{1}{\sqrt{n}}(K(\hat{\beta}) - K(\beta_0)) \in R^\varepsilon\right) \\ &+ \left|\mathbb{P}\left(\frac{1}{\sqrt{n}}(g_n(\beta_0) - g(\beta_0)) \in A^\varepsilon\right) - \mathbb{P}(N(0, \Gamma) \in A^\varepsilon)\right| \\ &+ |\mathbb{P}(N(0, \Gamma) \in A^\varepsilon) - \mathbb{P}(N(0, \Gamma) \in A)|, \end{aligned}$$

where $A^\varepsilon = \{x \in \mathbb{R}^p : \|x - y\| \leq \varepsilon, \forall y \in A\}$ and $R^\varepsilon = \{x \in \mathbb{R}^p : |x_j| \geq \varepsilon \text{ for some } j\}$. By Barry-Esseen bound (Raič [2019]), we have

$$\sup_{A \in \mathcal{A}} \left| \mathbb{P}\left(\frac{1}{\sqrt{n}}(g_n(\beta_0) - g(\beta_0)) \in A^\varepsilon\right) - \mathbb{P}(N(0, \Gamma) \in A^\varepsilon) \right| \leq Cn^{-1/2}. \quad (\text{C.8})$$

By Nazarov's inequality (Nazarov [2003]), we have

$$\sup_{A \in \mathcal{A}} |\mathbb{P}(N(0, \Gamma) \in A^\varepsilon) - \mathbb{P}(N(0, \Gamma) \in A)| \leq C\varepsilon. \quad (\text{C.9})$$

Choose $\varepsilon = n^{v-1/6}$ for any $v > 0$. By lemma C.2.2, we have

$$\mathbb{P}(|\hat{\beta} - \beta_0| > n^{-1/3}) \leq Cn^{-1/3}.$$

Combining lemma 2, let $u = \delta = n^{-1/3}$ and $t = n^{v+1/3}$. Then, we have

$$\mathbb{P}(|K(\hat{\beta}) - K(\beta)| \geq n^{v+1/3}) \leq c_1(n^{3p/2}e^{-c_2n^v} + n^{-1}),$$

which together with (C.8), (C.9) leads to

$$|\mathbb{P}(n^{-1/2}(g_n(\hat{\beta}) - g(\hat{\beta})) \in A) - \mathbb{P}(N(0, \Gamma) \in A)| \leq Cn^{v-1/6}$$

Finally, by Taylor expansion and Bahadur representation for $\hat{\beta}$, we have

$$\begin{aligned} g_n(\hat{\beta}) - g(\hat{\beta}) &= n\nabla g(\beta_0)(\hat{\beta} - \beta_0) + g_n(\hat{\beta}) - g(\beta_0) + O(|\hat{\beta} - \beta_0|^2) \\ &= n\nabla g(\beta_0)(\hat{\beta} - \beta_0) + O_p(1/n), \end{aligned}$$

which aligned with previous bounds completes the proof of the first part. $\square$

Proof of Corollary 4.2.1.1: Choose $B = \{x \in \mathbb{R}^p : |x_1| \geq \sqrt{c \log n}\}$ and $A = \nabla g(\beta_0)B$.

Then

$$\mathbb{P}\left(\sqrt{n}\nabla g(\beta_0)(\hat{\beta} - \beta_0) \in A\right) = \mathbb{P}(\sqrt{n}|\hat{\beta}_1 - \beta_{0,1}| \geq \sqrt{c \log n})$$

and with condition $0 \leq c < \sigma_{11}^2/6$, we have

$$\mathbb{P}(N(0, \Gamma) \in A) = \mathbb{P}\left(N(0, \tilde{\Gamma}) \in B\right) = \mathbb{P}\left(|N(0, \sigma_{11}^2)| \geq \sqrt{c \log n}\right) \asymp n^{-c/\sigma_{11}^2}.$$

So $A \in \mathcal{A}$ and applying Theorem 4.2.1 with $v = 1/12 - c/(2\sigma_{11}^2)$, we can prove the result. $\square$