Three Essays on Extremal Quantiles
by
Yichong Zhang
Department of Economics
Duke University

Date: ____________________
Approved:

__________________________
Shakeeb Khan, Co-supervisor

__________________________
Arnaud Maurel, Co-supervisor

__________________________
Federico Bugni

__________________________
Matt Masten

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics in the Graduate School of Duke University
2016
Abstract

Three Essays on Extremal Quantiles

by

Yichong Zhang

Department of Economics
Duke University

Date: __________________________

Approved:

______________________________
Shakeeb Khan, Co-supervisor

______________________________
Arnaud Maurel, Co-supervisor

______________________________
Federico Bugni

______________________________
Matt Masten

An abstract of a dissertation submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2016
Abstract

Extremal quantile index is a concept that the quantile index will drift to zero (or one) as the sample size increases. The three chapters of my dissertation consists of three applications of this concept in three distinct econometric problems. In Chapter 2, I use the concept of extremal quantile index to derive new asymptotic properties and inference method for quantile treatment effect estimators when the quantile index of interest is close to zero. In Chapter 3, I rely on the concept of extremal quantile index to achieve identification at infinity of the sample selection models and propose a new inference method. Last, in Chapter 4, I use the concept of extremal quantile index to define an asymptotic trimming scheme which can be used to control the convergence rate of the estimator of the intercept of binary response models.
To my family.
Contents

Abstract iv
List of Tables xi
List of Figures xiv
List of Abbreviations and Symbols xvi
Acknowledgements xvii
1 Introduction 1

2 Extremal Quantile Treatment Effects 3
  2.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3
  2.2 Definition, extreme value theory, and notation . . . . . . . . . . . . . 7
  2.3 Intermediate quantile treatment effects . . . . . . . . . . . . . . . . . 9
    2.3.1 The main result . . . . . . . . . . . . . . . . . . . . . . . . . . 9
    2.3.2 Estimation of the extreme value index . . . . . . . . . . . . . 18
  2.4 Extreme quantile treatment effects . . . . . . . . . . . . . . . . . . . 20
    2.4.1 The main result . . . . . . . . . . . . . . . . . . . . . . . . . . 20
    2.4.2 Asymptotic distribution under various boundary conditions . . 27
    2.4.3 Feasible normalizing factor . . . . . . . . . . . . . . . . . . . . 35
  2.5 Inference . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 38
    2.5.1 Inconsistency of the standard bootstrap inference method . . . 38
    2.5.2 Consistency of the b out of n bootstrap inference . . . . . . . 40
2.5.3 A robust confidence interval ........................................ 45
2.5.4 Inference theory for the 0-th QTE ............................... 48
2.6 Simulations .............................................................. 50
  2.6.1 Limiting distributions ............................................. 50
  2.6.2 Inference for the extreme QTE ............................... 51
  2.6.3 The robust confidence interval ............................... 56
  2.6.4 Inference for the 0-th QTE ............................... 57
2.7 Empirical application .................................................. 59
  2.7.1 Effect of maternal status on extremely low birth weights ... 59
  2.7.2 Effect of minority status on college preparation index ...... 63
2.8 Conclusion .............................................................. 68

3 Extremal Quantile Regressions for Selection Models and the Black-
White Wage Gap ......................................................... 70
  3.1 Introduction ........................................................... 70
  3.2 The set-up and nonparametric identification .................. 76
    3.2.1 Model and main result ....................................... 76
    3.2.2 The independence at infinity condition ................... 81
  3.3 Semiparametric estimation ......................................... 82
    3.3.1 Definition of the estimators ............................... 82
    3.3.2 Asymptotic properties and inference ..................... 87
    3.3.3 Choice of the quantile index ............................... 95
  3.4 Simulations ........................................................... 97
  3.5 Application to the black-white wage gap ....................... 103
    3.5.1 Evidence from the NLSY79 .................................. 104
    3.5.2 Across-cohort evolution .................................... 107
  3.6 Concluding remarks .................................................. 111
4 $\sqrt{n}$-Consistency of the Intercept of a Binary Response Model Under Tail Restrictions 113

4.1 Introduction ................................................................. 113
4.2 Definition of the Semiparametric Estimator ......................... 115
4.3 Asymptotic Properties .................................................... 118
4.4 Extensions ................................................................. 127
4.5 Simulations ............................................................... 132
4.6 Conclusion ................................................................. 138

A Appendix for Chapter 2 140

A.1 Appendix ................................................................. 140
   A.1.1 Illustrative examples .............................................. 140
   A.1.2 Details of simulation designs ................................... 143
   A.1.3 Additional simulation results .................................... 144
   A.1.4 Proof of Theorem 2.3.1 ......................................... 163
   A.1.5 Proof of Theorem 2.3.3 ......................................... 180
   A.1.6 Proof of Theorem 2.3.4 ......................................... 183
   A.1.7 Proof of Theorem 2.4.1 ......................................... 184
   A.1.8 Proof of Corollary 2.4.2 ....................................... 192
   A.1.9 Proof of Corollary 2.4.4 ....................................... 194
   A.1.10 Proof of Theorem 2.4.2 ....................................... 196
   A.1.11 Proof of Proposition 2.4.5 .................................... 196
   A.1.12 Proof of Theorem 2.5.1 ....................................... 196
   A.1.13 Proof of Theorem 2.5.2 ....................................... 197
   A.1.14 Proof of Corollary 2.5.1 ....................................... 200
   A.1.15 Proof of Theorem 2.5.3 ....................................... 201
   A.1.16 Proof of Proposition 2.5.2 .................................... 202
Bibliography 307
Biography 317
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Coverage of 95% ( b ) out of ( n ) bootstrap CI, sample size = 5,000</td>
<td>52</td>
</tr>
<tr>
<td>2.2</td>
<td>Coverage of 95% ( n ) out of ( n ) bootstrap CI, sample size = 5,000</td>
<td>53</td>
</tr>
<tr>
<td>2.3</td>
<td>Coverage of 95% CI. Sample size is 5,000.</td>
<td>58</td>
</tr>
<tr>
<td>2.4</td>
<td>Extreme order unconditional QTE of smoking status.</td>
<td>63</td>
</tr>
<tr>
<td>2.5</td>
<td>Index gap across campus and initial major</td>
<td>66</td>
</tr>
<tr>
<td>2.6</td>
<td>Index gap across campus and initial major</td>
<td>68</td>
</tr>
<tr>
<td>3.1</td>
<td>Examples of copulas satisfying (3.2.4).</td>
<td>82</td>
</tr>
<tr>
<td>3.2</td>
<td>Examples of copulas leading to a polynomial rate of convergence.</td>
<td>92</td>
</tr>
<tr>
<td>3.3</td>
<td>Monte Carlo simulations</td>
<td>99</td>
</tr>
<tr>
<td>3.4</td>
<td>OLS and median log-wage regression results (NLSY79)</td>
<td>105</td>
</tr>
<tr>
<td>3.5</td>
<td>Extremal quantile regression results (NLSY79)</td>
<td>106</td>
</tr>
<tr>
<td>3.6</td>
<td>Extremal quantile and median regression results (NLSY79-NLSY97)</td>
<td>108</td>
</tr>
<tr>
<td>3.7</td>
<td>Black-white wage gap with age restriction and additional premarket factors</td>
<td>109</td>
</tr>
<tr>
<td>4.1</td>
<td>Bias and root-MSE</td>
<td>134</td>
</tr>
<tr>
<td>4.2</td>
<td>Bias and root-MSE</td>
<td>135</td>
</tr>
<tr>
<td>4.3</td>
<td>Bias and root-MSE</td>
<td>136</td>
</tr>
<tr>
<td>4.4</td>
<td>Bias and root-MSE</td>
<td>136</td>
</tr>
<tr>
<td>4.5</td>
<td>Bias and root-MSE</td>
<td>137</td>
</tr>
<tr>
<td>4.6</td>
<td>Bias and root-MSE</td>
<td>138</td>
</tr>
</tbody>
</table>
A.1 Simulation designs used in Section 4.5.  ........................................ 144
A.2 Coverage of 95% b out of n bootstrap CI, sample size = 300  .......... 145
A.3 Coverage of 95% n out of n bootstrap CI, sample size = 300  .......... 146
A.4 Bias of the median-unbiased estimator, sample size = 300. All values are inflated by 100.  ................................................................. 147
A.5 root-MSE of the median-unbiased estimator, sample size = 300. All values are inflated by 100.  ................................................................. 148
A.6 median-bias of the median-unbiased estimator, sample size = 300. All values are inflated by 100.  ................................................................. 148
A.7 MAE of the median-unbiased estimator, sample size = 300. All values are inflated by 100.  ................................................................. 149
A.8 Coverage of 95% CI, sample size = 300.  .............................................. 150
A.9 Bias of the median-unbiased 0-QTE estimator, sample size = 300. All values are inflated by 100.  ................................................................. 150
A.10 root-MSE of the median-unbiased 0-QTE estimator, sample size = 300. All values are inflated by 100.  ................................................................. 151
A.11 median-bias of the median-unbiased 0-QTE estimator, sample size = 300. All values are inflated by 100.  ................................................................. 151
A.12 MAE of the median-unbiased 0-QTE estimator, sample size = 300. All values are inflated by 100.  ................................................................. 151
A.13 Coverage of 95% b out of n bootstrap CI, sample size = 1,000 .... 152
A.14 Coverage of 95% n out of n bootstrap CI, sample size = 1,000 .... 152
A.15 Bias of the median-unbiased estimator, sample size = 1,000. All values are inflated by 1,000.  ................................................................. 154
A.16 root-MSE of the median-unbiased estimator, sample size = 1,000. All values are inflated by 100.  ................................................................. 154
A.17 median-bias of the median-unbiased estimator, sample size = 1,000. All values are inflated by 1,000.  ................................................................. 155
A.18 MAE of the median-unbiased estimator, sample size = 1,000. All values are inflated by 100.  ................................................................. 155

xii
# List of Figures

2.1  Categorization of the asymptotic distribution over the quantile index .......................... 4  
2.2  Finite minimizers ........................................................................................................ 29  
2.3  Continuum of minimizers ............................................................................................ 31  
2.4  Mixture of minimizers ................................................................................................ 34  
2.5  QQplot against EV law ................................................................................................. 51  
2.6  QQplot against Normal law ......................................................................................... 51  
2.7  Coverage across quantiles ............................................................................................ 55  
2.8  Coverage across subsample size .................................................................................. 56  
2.9  Coverage across quantiles ............................................................................................ 57  
2.10 Coverage across subsample size .................................................................................. 59  
3.1  Relationship between MSE (Y-axis) and $\tau_n$ (X-axis), $\hat{\Delta}^1$ ....................... 100  
3.2  Relationship between coverage (Y-axis) and $\tau_n$ (X-axis), $\hat{\Delta}^1$ ..................... 102  
A.1  Coverage across quantiles ............................................................................................ 146  
A.2  Coverage across subsample size .................................................................................. 147  
A.3  Coverage across quantiles ............................................................................................ 149  
A.4  Coverage across subsample size .................................................................................. 150  
A.5  Coverage across quantiles ............................................................................................ 153  
A.6  Coverage across subsample size .................................................................................. 153  
A.7  Coverage across quantiles ............................................................................................ 156
List of Abbreviations and Symbols

Abbreviations

QTE    Quantile Treatment Effect.
QTT    Quantile Treatment Effect on Treated.
CPI    College Preparation Index.
EV     Extreme Value.
Acknowledgements

I am deeply grateful to my co-advisors Shakeeb Khan and Arnaud Maurel, as well as my committee members Federico Bugni and Matt Masten for their guidance and encouragement. I also thank Xavier D’Haultfœuille, Jia Li, and participants in the microeconometrics lunch at Duke, UCSB, University of Connecticut, University of Warwick, University of Western Ontario, Singapore Management University, and the 11th World Congress of the Econometric Society for their comments. All remaining errors are mine.
Introduction

Extremal quantile index is a concept that the quantile index will drift to zero (or one) as the sample size increases. My dissertation applies this concept to three distinct econometrics problems. The first problem is how to conduct estimation and inference of extremal quantile treatment effects, which is considered in Chapter 2. In this solo work, I establish asymptotic theory and inference method for quantile treatment effect estimators when the quantile index is close or equal to zero. Such quantile treatment effects are of interest in many economic applications, such as the effect of maternal smoking on an infant’s adverse birth outcomes. When the quantile index is close to zero, the sparsity of data jeopardizes conventional asymptotic theory and bootstrap inference. When the quantile index is zero, there are no existing inference methods directly applicable in the treatment effect context. This paper establishes new estimation and inference theory for cases close or equal to zero. I illustrate the inference method with an application to estimate and test the effect of maternal smoking on the lower quantiles of infants’ birth weights.

The second problem is how to conduct inference method for sample selection
models without large support exclusive variables, which is considered in Chapter 3. This is a joint work with Xavier D’Haultfoeuille\textsuperscript{1} and Arnaud Maurel\textsuperscript{2}. In this work, we consider the estimation of a semiparametric location-scale model subject to endogenous selection, in the absence of an instrument or a large support regressor. Identification relies on the independence between the covariates and selection, for arbitrarily large values of the outcome. In this context, we propose a simple estimator, which combines extremal quantile regressions with minimum distance. We establish the asymptotic normality of this estimator by extending previous results on extremal quantile regressions to allow for selection. Finally, we apply our method to estimate the black-white wage gap among males from the NLSY79 and NLSY97. We find that premarket factors such as AFQT and family background characteristics play a key role in explaining the level and evolution of the black-white wage gap.

The last problem is when the convergence rate of the special-regressor estimator proposed by Lewbel (1997) and Lewbel (2000) is parametric, which is considered in Chapter 4. It is a joint work with Lili Tan\textsuperscript{3}. In this chapter, we consider the $\sqrt{n}$-consistency of the estimator of intercept of the binary response model when the supports of both the special regressor $V$ and the error term $\varepsilon$ are the whole real line. In this case, Khan and Tamer (2010) have shown that the convergence rate of the estimator depends on the relative thickness of the tails of $V$ and $\varepsilon$, but did not provide sufficient conditions for $\sqrt{n}$-consistency. This paper provides such sufficient conditions and an asymptotic trimming scheme that enables $\sqrt{n}$-consistency. It also provides sufficient conditions for non-existence of any $\sqrt{n}$-consistent estimator. Finally, we extend the sufficient conditions for $\sqrt{n}$-consistency into a full-blown model with endogenous regressors, as was first considered by Dong and Lewbel (2015).

\textsuperscript{1} CREST. E-mail address: xavier.dhaultfoeuille@ensae.fr.
\textsuperscript{2} Duke University, NBER and IZA. E-mail address: apm16@duke.edu.
\textsuperscript{3} Yunnan University. E-mail address: tanlili@ynu.edu.cn.
Extremal Quantile Treatment Effects

2.1 Introduction

Economic theory usually predicts that the sign and magnitude of treatment effects vary depending on one’s place in the overall distribution of outcomes. This heterogeneity can be captured by quantile treatment effects (QTEs). In many economic applications, the populations located at the tail of the outcome distribution, such as infants’ low birth weights or students’ low scores are of particular interest. In this case, however, researchers encounter both the usual missing counterfactual and data sparsity; that is, there are not many observations at the tails. While previous literature has considered the two problems separately, it is still an open question how to cope with both simultaneously while conducting proper statistical inferences.

This paper addresses both issues simultaneously. I establish new asymptotic theory and inference method for an estimator of the QTE for low-rank populations. To deal with the usual missing counterfactual problem, I assume unconfoundedness and rely on the propensity score to identify QTEs. To address the data sparsity, I model a small quantile index $\tau$ as a drifting object with sample size $n$; that is,
\[ \tau := \tau_n \to 0 \text{ as } n \to \infty. \] Then I use the device of extremal quantiles to derive a new asymptotic approximation for the finite sample distribution of the QTE estimator when the quantile index \( \tau \) is close to zero.

I establish the asymptotic properties for extremal QTE estimators when \( \tau_n \to 0 \). I find that there are two asymptotic distributions of the estimator of \( \tau_n \)-th QTE, depending on how fast \( \tau_n \) approaches zero. Following the terminology used in Chernozhukov (2005a), I say \( \tau_n \) is intermediate when \( \tau_n \to 0 \) and \( \tau_n n \to \infty \). In this case, I show that the asymptotic distribution for the proposed estimator of QTE is still Gaussian. Again, following Chernozhukov (2005a), when \( \tau_n \to 0, \tau_n n \to k \), for some \( k > 0 \), I say \( \tau_n \) is extreme. In this case, I show that the asymptotic distribution is non-Gaussian. Figure 2.1 summarizes this of the quantile index.

![Figure 2.1: Categorization of the asymptotic distribution over the quantile index](image)

For inference, when the quantile index is intermediate, I show that the standard bootstrap confidence interval (CI) for the QTE estimator is consistent. For the extreme-order quantile case, I first prove that the conventional bootstrap CI does not control size. I then propose a resampling method that is uniformly consistent over a range of quantile indices. Lastly, by considering a linear combination of extreme QTE estimators with carefully chosen weights, I construct a consistent CI for the 0-th QTE without imposing additional restrictions or extrapolations.

I propose a quantile-order-category-selection procedure similar to the identification-category-selection procedure used in Andrews and Cheng (2012). The difference here is I have two thresholds while they only have one. When the quantile index is
smaller than the first threshold, the extreme-order quantile asymptotic distribution is expected to approximate the finite sample distribution of the QTE estimator better than the normal approximation. In this case, I suggest researchers use the new resampling CI developed in this paper to conduct inference. In simulation, I examine the performance of this threshold in 16 simulation designs with small, moderate, and large size samples. In all cases, I find that when the criterion is satisfied, the new resampling CI controls size while the standard bootstrap CI undercovers (over-rejects) by as much as 18 absolute percentage points. When the quantile index is greater than the second threshold, I prove that the standard bootstrap CI is consistent. Last, when the quantile index is in between the first and second threshold, I construct a robust CI which is conservative.

My resampling inference method gives empirical researchers tools to estimate, infer, and test QTEs for low-rank populations. This method can be used in a number of economics applications. For instance, when focusing on the students body who are admitted by the university, the college preparation index of low-rank students reflects the tolerance of low academic performance in the college’s admission policy. My methods allow researchers to estimate the college preparation index gap between low-scoring minority and non-minority students while controlling for family background. Researchers can use this gap as a measure of racial preference in colleges’ admission selections policies. In another example, the extremely low or lower boundary of babies’ birth weights represents the severity of adverse birth outcomes, which have been found to result in large economic costs. See Abrevaya (2002) for a detailed discussion. My methods allow researchers to make inferences about the effect of maternal smoking on the lower tail of babies’ birth weights distribution.

My paper addresses the problem of missing counterfactual and data sparsity simultaneously. I build on the previous literature that address only one issue at
a time. For the treatment effect literature addressing the missing counterfactual problem, I adapt the same unconfoundedness assumption as Bitler et al. (2006), Chernozhukov et al. (2013a), Firpo (2007), and Hirano et al. (2003). For further applications of QTEs, see Card (1996) and DiNardo et al. (1996), for example.

For the extremal quantile literature addressing the data sparsity problem, Chernozhukov (2005a), Chernozhukov and Fernández-Val (2011), Feigin and Resnick (1994), Knight (2001), Portnoy and Jurečková (1999), and Smith (1994) assume that the conditional quantile is linear. In particular, the extremal QTE considered in this paper is closely related to the linear extremal quantile regression (LEQR) investigated in Chernozhukov (2005a) and Chernozhukov and Fernández-Val (2011), but substantially differs in two aspects. First, the QTE considered in this paper has a causal interpretation by addressing the problem of missing counterfactuals, while the causal interpretation for the coefficient in the LEQR relies on the assumption that the treatment variable is exogenous at the tails. Second, since the QTE is an unconditional object, I do not assume the linearity of the conditional quantiles of $(Y_1, Y_0)$ given $X$.

The literature on extremal percentiles also addresses the data sparsity problem. See, for example, Bertail et al. (2004), Bickel and Sakov (2008), and Dekkers and De Haan (1989). The key difference between these papers and mine is that I include additional covariates $X$ and use propensity score $P(X)$ to correct the selection bias.

Last, my paper is related to the concept of drifting sequence asymptotics. This concept goes back to Pitman (1949) using Pitman drift to characterize power functions. Recently, the concept has been used in the context of weak instruments by, for example, Stock J (2008), Stock and Yogo (2005), and other various models by Andrews and Cheng (2012), Andrews and Cheng (2013), Chen et al. (2014), and Khan and Nekipelov (2013).
The rest of the paper is organized as follows. Section 2 defines the parameters of interest, introduces additional notation, and provides relevant background on extreme value theory. Section 3 considers the asymptotic properties of the estimator for intermediate QTEs. Section 4 considers the asymptotic properties of the estimator for extreme QTEs. Section 5 establishes the inference theory for the estimator of extreme as well as the 0-th QTE and provides a step-by-step description of implementation. Section 6 contains a simulation study. Section 7 applies the method to examine the effect of maternal smoking on infants’ birth weights. Section A.1 in the Appendix computes the limiting distributions for the extreme-order quantile estimators under various boundary conditions. Section A.2 provides details of simulation designs used in Section 4.5. Section A.3 contains additional simulation results. The rest of the Appendix provides proofs.

2.2 Definition, extreme value theory, and notation

This section defines QTEs and quantile treatment effects on the treated (QTT’s), which are the parameters I focus on. It also introduces some extreme value theory, especially the meaning of domain of attraction and the extreme value (EV) index, which will be used throughout the paper. Last, it provides the detail of two weak convergence concepts I rely on.

Denote the outcomes for treated and control groups as $Y_1$ and $Y_0$, respectively. The treatment status is denoted as $D$, where $D = 1$ means treated and $D = 0$ means untreated. The econometrician can only observe $(Y, X, D)$ where $Y = Y_1 D + Y_0 (1 - D)$, and $X$ is a collection of confounders. The propensity score $P(D = 1 | X = x)$ is denoted as $P(x)$. The parameters of interest are the $\tau$-th QTE defined as

$$q(\tau) := q_1(\tau) - q_0(\tau)$$
and the $\tau$-th QTT defined as

$$ q_{D=1}(\tau) := q_{1|D=1}(\tau) - q_{0|D=1}(\tau), $$

in which $q_{j}(\tau)$ and $q_{j|D=1}(\tau)$ denote the $\tau$-th quantile of random variables $Y_j$ and $Y_j|D = 1$, respectively.

Next, I introduce some extreme value theory, which will be used when I characterize the asymptotic theories in Section 2.3 and 2.4. The cumulative distribution function (c.d.f.) $F$ belongs to the domain of attraction of generalized extreme value distributions if there exist sequences $(\alpha_n)_{n \in \mathbb{N}}$, $(\beta_n)_{n \in \mathbb{N}}$ and a c.d.f. $G$ indexed by a parameter $\xi$, such that, for any independent draws $(U_1, ..., U_n)$ from $F$, $\alpha_n(\min(U_1, ..., U_n) - \beta_n)$ converges in distribution to $G$. Here, $F$ belongs to the domain of attraction of generalized extreme value distributions with a parameter $\xi$ called the extreme value (EV) index. Based on the value of $\xi$, $F$ has three types of tails. In particular, for

- type 1 tails ($\xi = 0$): as $z \to s_l$ \quad $F(z + va(z)) \sim F(z)e^v$, \quad $\forall v \in \mathbb{R}$,
- type 2 tails ($\xi > 0$): as $z \to s_l = -\infty$ \quad $F(vz) \sim v^{-1/\xi}F(z)$, \quad $\forall v > 0$,
- type 3 tails ($\xi < 0$): as $z \to s_l > -\infty$ \quad $F(vz) \sim v^{-1/\xi}F(z)$, \quad $\forall v > 0$,

where $a(z) := \int_{s_l}^z F(v)dv/F(z)$ for some $z > s_l$, $s_l$ is the lower end point of the support of $U$, for two generic functions $f_1(\cdot)$ and $f_2(\cdot)$, $f_1(z) \sim f_2(z)$ if

$$ \frac{f_1(z)}{f_2(z)} \to 1, \text{ as } z \to s_l. $$

For example, normal, T, and Beta distributions have type 1, 2, and 3 tails, respectively.

Last, I provide two weak convergence concepts this paper will rely on. $U_n \rightsquigarrow U$ indicates weak convergence as defined by Van der Vaart and Wellner (1996). When
$U_n$ and $U$ are $k$-dimensional elements, the space of the sample path is $\mathbb{R}^k$ equipped with the Euclidean metric. When $U_n$ and $U$ are stochastic processes, the space of the sample path will be specified later in each different context. For this paper, the space is either $l^\infty(\{v \in \mathbb{R} : |v| < B\})$, for some positive $B$ equipped with the sup norm or the Skorohod space $\mathcal{D}([-B, B])$, for some positive $B$ equipped with the Skorohod metric$^1$.

2.3 Intermediate quantile treatment effects

In Theorems 2.3.1 and 2.3.2, this section establishes the asymptotic theory for $\tau_n$-th QTE when $\tau_n$ is intermediate. These theorems give the first main theoretical result of the paper: that the asymptotic distribution of the estimator of an intermediate QTE is still Gaussian as in the regular quantile index case. The asymptotic theory established here can be used to construct a uniform confidence band for both intermediate and extreme QTE, to estimate the EV index (which is analyzed in detail in Section 2.3.2), and to deal with the sample selection problem as in D’Haultfoeuille et al. (2015).

2.3.1 The main result

This section first establishes the asymptotic normality for the $\tau_n$-th quantile estimators $\hat{q}_j(\tau_n), j = 0, 1$, when $\tau_n$ is intermediate. Theorem 2.3.2 takes this result one step further and establishes the asymptotic normality of $\tau_n$-th QTE estimator $\hat{q}_1(\tau_n) - \hat{q}_0(\tau_n)$. Then I discuss the two inference methods for the intermediate QTE: the analytical method and the bootstrap.

Recall the setup in Section 2.2. I further assume:

Assumption 1.

$^1$ To differentiate, $D$ is reserved for the binary treatment status and $\{D_{i,j}\}_{i=1}^\infty, j = 0, 1$ are the sets of random variables defined in the limiting objective function in Section 2.4.
(1) (random sample): \( \{Y_i, D_i, X_i\}_{i=1}^n \) is i.i.d.

(2) (unconfoundedness): \( (Y_1, Y_0) \perp D |X \).

(3) (common support): \( \text{Supp}(X) \), the support of \( X \), is compact. For some \( c > 0 \), \( c < P(x) < 1 - c \), \( \forall x \in \text{Supp}(X) \).

The unconfoundedness assumption states that the potential outcomes are independent of the treatment status conditional on additional covariates \( X \). Although this is a strong assumption, it has been widely used in both theoretical investigations and empirical studies. See, for example, Bitler et al. (2006), Chernozhukov et al. (2013a), Firpo (2007), Hirano et al. (2003), Rosenbaum and Rubin (1983), and the references therein. For extremal QTEs, it is natural to first start with the simple unconfoundedness condition. When the quantile index is regular, that is, bounded away from 0 and 1, papers such as Abadie et al. (2002), Chernozhukov and Hansen (2005), Chernozhukov and Hansen (2008), and Frölich and Melly (2013) extend the assumption to allow for endogenous treatment status and rely on an instrumental variable to correct the selection bias. Similar strategies can be applied here to the extremal quantile case. While important, I leave the problem of establishing the corresponding asymptotic theory to future research.

**Assumption 2.** \( \tau_n \) is intermediate. This is,

(1) \( \tau_n \to 0 \) as \( n \to \infty \).

(2) \( \tau_n n \to \infty \) as \( n \to \infty \).

I define \( \hat{q}(\tau_n) \), the estimator of the \( \tau_n \)-th QTE, as \( \hat{q}(\tau_n) := \hat{q}_1(\tau_n) - \hat{q}_0(\tau_n) \) and \( \hat{q}_{D=1}(\tau_n) \), the estimator of \( \tau_n \)-th QTT, as \( \hat{q}_{D=1}(\tau_n) := \hat{q}_{1\|D=1}(\tau_n) - \hat{q}_{0\|D=1}(\tau_n) \). Under Assumption 1, Firpo (2007) pointed out that the the four quantiles \( q_1(\tau) \), \( q_0(\tau) \), \( q_{1\|D=1}(\tau) \), and \( q_{0\|D=1}(\tau) \) for any \( \tau \in (0,1) \) are identified based on the following four
moment equalities:

\[
\mathbb{E} \left[ \frac{D}{P(X)} \left( \tau - 1 \{ Y \leq q_1(\tau) \} \right) \right] = 0, \quad \mathbb{E} \left[ \left( \frac{1 - D}{1 - P(X)} \right) \left( \tau - 1 \{ Y \leq q_0(\tau) \} \right) \right] = 0,
\]

and

\[
\mathbb{E} \left[ D(\tau - 1 \{ Y \leq q_{\|D=1}(\tau) \}) \right] = 0, \quad \mathbb{E} \left[ \frac{(1 - D)P(X)}{1 - P(X)} \left( \tau - 1 \{ Y \leq q_{\|D=1}(\tau) \} \right) \right] = 0,
\]

respectively.

Therefore, despite the extremal feature of the quantile index, the natural sample estimator \( \hat{q}_1(\tau_n) \) for the \( \tau_n \)-th quantile of \( Y_1 \) can be computed through an inverse propensity score weighted quantile regression:

\[
\hat{q}_1(\tau_n) := \arg \min_{q \in \mathbb{R}} \sum_{i=1}^{n} \frac{D_i}{\hat{P}(X_i)} (Y_i - q) (\tau_n - 1 \{ Y_i \leq q \}). \tag{2.3.1}
\]

Similarly, \( \hat{q}_0(\tau_n) \), an estimator of the \( \tau_n \)-th quantile of \( Y_0 \), can be computed as

\[
\hat{q}_0(\tau_n) := \arg \min_{q \in \mathbb{R}} \sum_{i=1}^{n} \frac{1 - D_i}{1 - \hat{P}(X_i)} (Y_i - q) (\tau_n - 1 \{ Y_i \leq q \}). \tag{2.3.2}
\]

For estimating the QTT, \( \hat{q}_{\|D=1}(\tau_n) \) and \( \hat{q}_{\|D=1}(\tau_n) \) can be computed as

\[
\hat{q}_{\|D=1}(\tau_n) := \arg \min_{q \in \mathbb{R}} \sum_{i=1}^{n} \frac{D_i}{n \sum_{i=1}^{n} D_i} (Y_i - q) (\tau_n - 1 \{ Y_i \leq q \}),
\]

and

\[
\hat{q}_{\|D=1}(\tau_n) := \arg \min_{q \in \mathbb{R}} \sum_{i=1}^{n} \frac{1 - D_i}{n \sum_{i=1}^{n} D_i} \frac{\hat{P}(X_i)}{1 - \hat{P}(X_i)} (Y_i - q) (\tau_n - 1 \{ Y_i \leq q \}).
\]

Following Firpo (2007) and Hirano et al. (2003), \( \hat{P}(X) \), the propensity score, is estimated by the sieve method of fitting a series logistic model. I denote the logistic
c.d.f. by $L(a) := \exp(a)/(1 + \exp(a))$. $H_h(x) := (r_{1h}(x), \ldots, r_{hh}(x))'$ is a $h$-vector of power series of $x$. Then $\hat{P}(x) := L(H_h(x)'\hat{\pi}_h)$ with

$$\hat{\pi}_h := \arg\max_{\pi \in \mathbb{R}^h} \sum_{i=1}^n (D_i \log L(H_h(X_i)'\pi) + (1 - D_i) \log(1 - L(H_h(X_i)'\pi))).$$

For brevity, the rest of the paper only considers the estimation of $\hat{q}_1(\tau_n)$, $\hat{q}_0(\tau_n)$, and $\hat{q}(\tau_n)$. The asymptotic results for $\hat{q}_{D=1}(\tau_n)$, $\hat{q}_{D=0}(\tau_n)$, and $\hat{q}_{D=1}(\tau_n)$ can be derived in a similar manner.

Furthermore, instead of only one quantile index $\tau_n$, I focus on a range of them. That is, $k\tau_n$, $k \in [\kappa_1, \kappa_2]$ for some fixed and known constants $\kappa_1$ and $\kappa_2$ such that $0 < \kappa_1 < \kappa_2 < \infty$. This is because I will derive a uniform asymptotic theory for the process $\{ (\hat{q}_1(k\tau_n), \hat{q}_0(k\tau_n)) : k \in [\kappa_1, \kappa_2] \}$. For each $k$, $\hat{q}(k\tau_n) := \hat{q}_1(k\tau_n) - \hat{q}_0(k\tau_n)$

where

$$\hat{q}_1(k\tau_n) := \arg\min_{q \in \mathbb{R}} \sum_{i=1}^n \frac{D_i}{\hat{P}(X_i)} (Y_i - q)(k\tau_n - 1\{Y_i \leq q\})$$

and

$$\hat{q}_0(k\tau_n) := \arg\min_{q \in \mathbb{R}} \sum_{i=1}^n \frac{1 - D_i}{1 - \hat{P}(X_i)} (Y_i - q)(k\tau_n - 1\{Y_i \leq q\}).$$

The following sufficient regularity conditions are adapted from Assumptions A.1 and A.2 of Firpo (2007):

**Assumption 3.**

1. The density of $X$ is bounded above and bounded away from 0 over its support.
2. The propensity score $P(x)$ is $s$-times continuously differentiable with all the derivatives bounded.
3. $E(k\tau_n - 1\{Y_j \leq q_j(k\tau_n)\}|x)$ is $t$-times continuously differentiable in $x$ with all
derivatives bounded by $M_n$ uniformly over $(x, k) \in \text{Supp}(X) \times [k_1, k_2]$.

(4) The order of the series is $h = CN^c$ for some constants $C$ and $c$ such that $c < \frac{1}{6}$, 
$\tau_n n^{1+c(6-\frac{c}{2})} \to 0$, $\frac{M_n n^{1-k}}{\tau_n} \to 0$, and $n^{1/5-1/\tau_n} \to 0$, where $r$ is the dimension of $X$.

Assumption 3(1) and 3(2) are common in the sieve estimation literature. Assumption 3(3) and 3(4) are tailored to fit the special case in which the quantile index is intermediate and the derivative of the quantile varies with the sample size. In fact, the magnitude of $M_n$ depends on the tail behavior of $Y_j$ conditional on $X$. When the density of $Y_j \mid X$ vanishes on its lower tail, $M_n$ decreases to zero. When the density of $Y_j \mid X$ diverges on its lower tail (such as a beta distribution with the first shape parameter less than 1), $M_n$ diverges to infinity. Last, Assumption 3(3) and 3(4) can be further relaxed by using the double robust method as illustrated in Firpo and Rothe (2014).

Next, I impose regularity conditions on the tails of $Y_1$ and $Y_0$.

**Assumption 4.** For $j = 0, 1$

(1) $Y_j, Y_j \mid X$ are continuously distributed with density $f_j(\cdot)$ and $f_j(\cdot \mid X)$.

(2) $f_j(\cdot)$ is monotone at its lower tails.

(3) The c.d.f. of $Y_j$ belongs to the domain of attraction of generalized EV distributions with the EV index $\xi_j$.

These restrictions are mild. Assumption 20(1) is common in quantile regression literature. Assumption 20(2) refers to the tail of the distribution, which is satisfied by most well-known distributions. Assumption 20(3) is a standard condition in extreme value theory and is satisfied by almost all continuous distributions.

Before stating the first main theoretical result of the paper, I introduce the normalizing factor $\lambda_{j,n}(k)$ for $\hat{q}_j(k \tau_n)$:

$$
\lambda_{j,n}(k) := \sqrt{\frac{n}{k \tau_n}} f_j(q_j(k \tau_n)) \quad \text{for} \quad j = 0, 1 \quad \text{and} \quad k \in [k_1, k_2]. \quad (2.3.3)
$$
Recall that for the regular quantile estimation, the convergence rate is $\sqrt{n}$ and the asymptotic variance is $\frac{\tau(1-\tau)}{f_j^2(q_j(\tau))}$. By moving the asymptotic standard deviation to the same side of the convergence rate, we obtain a normalizing factor

$$\sqrt{\frac{n}{\tau(1-\tau)}} f_j(q_j(\tau)).$$

Then letting $\tau := \tau_n \to 0$, we heuristically obtain the normalizing factor for the intermediate-order quantile estimators defined in (2.3.3).

**Theorem 2.3.1.** If Assumptions 1–20 hold, then

$$\left( \lambda_{1,n}(k)(\hat{q}_1(k\tau_n) - q_1(k\tau_n)), \lambda_{0,n}(k)(\hat{q}_0(k\tau_n) - q_0(k\tau_n)) \right)$$

as a two-dimensional stochastic process indexed by $k$ is asymptotically tight under the uniform metric.

In addition, if there exist functions $H_1(k_1,k_2)$, $H_0(k_1,k_2)$, and $H_{10}(k_1,k_2)$ on $(k_1, k_2) \in [\kappa_1, \kappa_2] \times [\kappa_1, \kappa_2]$ such that, as $\tau_n \to 0$,

$$\frac{1}{\tau_n} \mathbb{E} \left[ \frac{P(Y_1 \leq q_1(\min(k_1, k_2)\tau_n)|X)}{P(X)} - \frac{P(X)}{1 - P(X)} P(Y_1 \leq q_1(k_1\tau_n)|X) P(Y_1 \leq q_1(k_2\tau_n)|X) \right] \to H_1(k_1,k_2)$$

$$\frac{1}{\tau_n} \mathbb{E} \left[ \frac{P(Y_0 \leq q_0(\min(k_1, k_2)\tau_n)|X)}{1 - P(X)} - \frac{P(X)}{1 - P(X)} P(Y_0 \leq q_0(k_1\tau_n)|X) P(Y_0 \leq q_0(k_2\tau_n)|X) \right] \to H_0(k_1,k_2)$$

$$\frac{1}{\tau_n} \mathbb{E} P(Y_0 \leq q_0(k_1\tau_n)|X) P(Y_0 \leq q_0(k_2\tau_n)|X) \to H_{10}(k_1,k_2),$$

then for $k \in [\kappa_1, \kappa_2]$,

$$\left( \lambda_{1,n}(k)(\hat{q}_1(k\tau_n) - q_1(k\tau_n)), \lambda_{0,n}(k)(\hat{q}_0(k\tau_n) - q_0(k\tau_n)) \right) \rightsquigarrow B(k)$$
where \( \mathcal{B}(k) \) is a Brownian bridge with covariance kernel

\[
\mathcal{H}(k_1, k_2) := \begin{pmatrix}
H_1(k_1, k_2) & H_{1,0}(k_1, k_2) \\
\sqrt{k_1 k_2} & H_{1,0}(k_1, k_2)
\end{pmatrix}.
\]

This theorem shows that the asymptotic distribution of the intermediate QTE estimator is still Gaussian, just as when the quantile index is fixed. Intuitively, this is because for \( j = 0, 1, \hat{q}_j(\tau_n) \) can be interpreted as a cutoff for which the number of \( \{Y_{i,j}\}_{i=1}^n \) below and above the cutoff are of the same order of \( n\tau_n \) and \( n(1 - \tau_n) \), respectively. When \( \tau_n \) is intermediate, both orders diverge to infinity, which is the same as the case in which \( \tau \) is regular, that is, strictly between zero and one. Thus the shapes of asymptotic distributions under regular and intermediate-order quantile indices are the same.

The difference between the regular and intermediate-order quantile asymptotic properties is that for the intermediate case, there is no information gain through nonparametrically estimating the propensity score \( P(x) \). From the proof of Theorem 2.3.1, the influence function for \( \hat{q}_j \) is

\[
\phi_{i,1,n} := \frac{1}{\sqrt{\tau_n}} \left[ \frac{D_i}{P(X_i)} T_{i,1,n} - \frac{E(T_{i,1,n}|X_i)}{P(X_i)} (D_i - P(X_i)) \right]
\]

where

\[ T_{i,1,n} := \tau_n - \mathbb{1}\{Y_i \leq q_1(\tau_n)\}. \]

In \( \phi_{i,1,n} \), the second term

\[ \frac{\mathbb{E}(T_{i,1,n}|X_i)}{P(X_i)} (D_i - P(X_i)) \]

represents the information gain and is asymptotically negligible compared to the first term \( \frac{D_i}{P(X_i)} T_{i,1,n} \).
I next turn to the asymptotic theory of $\hat{q}(\tau_n) := \hat{q}_1(\tau_n) - \hat{q}_0(\tau_n)$. From Theorem 2.3.1, I can make two observations: (1) the normalizing factors proposed in Theorem 2.3.1 are not feasible, and (2) the tail behaviors of $Y_1$ and $Y_0$, and thus the convergence rates for $\hat{q}_1(\tau_n)$ and $\hat{q}_0(\tau_n)$, are not necessarily the same. For the first point, I follow Chernozhukov (2005a) to build a feasible normalizing factor based on quantile difference with spacing parameter $m > 1$. To address the second point, I use the following assumption.

Assumption 5.

$$\frac{q_1(m\tau_n) - q_1(\tau_n)}{q_0(m\tau_n) - q_0(\tau_n)} \to \rho \in [0, +\infty].$$

Assumption 5 aims to bridge the normalizing factors of $\hat{q}_1(\tau_n)$ and $\hat{q}_0(\tau_n)$ by $\rho$. When $\rho = 0$, the convergence rate for $\hat{q}_0$ is slower so the estimation error of $\hat{q}_1(\tau_n)$ is asymptotically negligible. On the other hand, if $\rho = \infty$, $\hat{q}_0(\tau_n)$ is super-consistent compared to $\hat{q}_1(\tau_n)$ and thus can be treated as known. Last, when $\rho \in (0, \infty)$, the convergence rates for $\hat{q}_1(\tau_n)$ and $\hat{q}_0(\tau_n)$ are the same. For analytical inference, when $\tau_n$ is intermediate, $\rho$ can be estimated by

$$\hat{\rho} = \frac{\hat{q}_1(m\tau_n) - \hat{q}_1(\tau_n)}{\hat{q}_0(m\tau_n) - \hat{q}_0(\tau_n)}.$$

Under Assumption 5, I define the feasible normalizing factor for $\hat{q}(\tau_n)$ as

$$\hat{\lambda}_n := \frac{\sqrt{n\tau_n}}{\max\left\{ (\hat{q}_1(m\tau_n) - \hat{q}_1(\tau_n)), (\hat{q}_0(m\tau_n) - \hat{q}_0(\tau_n)) \right\}}.$$

The next theorem shows that the intermediate QTE estimator is asymptotically normal with the feasible normalizing factor $\hat{\lambda}_n$. 

16
Theorem 2.3.2. Let \( C_1(\rho, m) := \left( \frac{1-m^{-\xi_1}}{\xi_1} \right)^{-1} \frac{\rho}{\max(1, \rho)} \), \( C_0(\rho, m) := \left( \frac{1-m^{-\xi_0}}{\xi_0} \right)^{-1} \frac{1}{\max(\rho, 1)}^2 \), and

\[
\Sigma_n := \text{Var}(C_1(\rho, m) \phi_{i,1,n} - C_0(\rho, m) \phi_{0,n,i})/\tau_n.
\]

If Assumptions 1–5 hold, then

\[
\Sigma_n^{-1/2} \hat{\lambda}_n(\hat{q}(\tau_n) - q(\tau_n)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).
\]

The additional assumptions in Theorem 2.3.1 are not needed. This is because the variance \( \Sigma_n \) needs not to be convergent to apply the central limit theorem. Based on Theorem 2.3.2, I can conduct inference by estimating \( \Sigma_n \) and referring to the standard normal critical value.

In addition, the next theorem shows that the standard bootstrap inference for the intermediate QTE is consistent. Let \( \hat{q}^*(\tau_n) \) be the estimator from the bootstrap sample and \( \tilde{C}_{a}(\tau_n) \) be the \( a \)-th quantile of \( \hat{q}^*(\tau_n) - \hat{q}(\tau_n) \) conditional on data. The two-sided \( 1 - a \)-th bootstrap CI for any \( a \in (0, 1) \) can be written as

\[
\text{CI}^{\text{boot}}(\tau_n) = \left( \hat{q}(\tau_n) - \tilde{C}_{a/2}(\tau_n), \hat{q}(\tau_n) - \tilde{C}_{1-a/2}(\tau_n) \right).
\]

Theorem 2.3.3. If Assumptions 1–5 hold, then

\[
\lim_{n \to \infty} P(q(\tau_n) \in \text{CI}^{\text{boot}}(\tau_n)) = 1 - a.
\]

For the intermediate-order percentiles, Falk (1991) has proven the validity of bootstrap inference. For the regression case, Chernozhukov (2000) points out that

\[\text{Here I adapt the convention that } \frac{c}{0} = 0, \text{ sign}(c) = 0 \text{ for any real number } c, \text{ and } \frac{1-m^{-\xi}}{\xi} = \log(m) \text{ when } \xi = 0.\]
the bootstrap inference is valid for linear intermediate-order quantile regressions. Recently, D’Haultfoeuille et al. (2015) proves that the bootstrap inference for intermediate-order quantile regression is valid in sample selection models. Here, I show that the bootstrap inference is also valid for the intermediate-order QTE estimator.

2.3.2 Estimation of the extreme value index

In this section, I focus on the estimation of EV indices $\xi_j$ for $j = 0, 1$. A consistent estimator of the EV index will be used in Section 2.5.4 to construct a consistent CI for the 0-th QTE. The result is also of independent interest because it contributes to the statistics literature on estimating the EV index when the data are missing randomly conditional on covariates. In the previous literature, attentions are paid on estimating the EV index for the observable $Y$. See Chapter 4 of Resnick (2007) for a textbook treatment on this topic. Alternatively, here the potential outcomes $(Y_1, Y_0)$ are not fully observed and the counterfactuals are only conditionally missing at random. Theorem 2.3.4 addresses this issue, proposes estimators of the EV indices for $Y_1$ and $Y_0$, and establishes their asymptotic properties.

The proposed EV index estimator follows the Pickands type as described in Section 4.5 of Resnick (2007). For some positive integer $R$, $\{w_r\}_{r=1}^R$ is a set of weights which sum to one. I estimate $\xi_j$, the EV index of $Y_j$, for $j = 0, 1$ by

$$\hat{\xi}_j := \sum_{r=1}^{R} \frac{w_r}{\log(l)} \log \left( \frac{\hat{q}_j(ml^r\tau_n) - \hat{q}_j(l^r\tau_n)}{\hat{q}_j(ml^{r-1}\tau_n) - \hat{q}_j(l^{r-1}\tau_n)} \right),$$

in which $l$ is some positive constant and $\tau_n$ is intermediate.

The intuition for the estimator is straightforward. If $Y_j$ has EV index $\xi_j$, $q_j(\tau)$ behaves as $\tau^{-\xi_j}$ as $\tau \to 0$. Then

$$\log \left( \frac{q_j(ml^r\tau_n) - q_j(l^r\tau_n)}{q_j(ml^{r-1}\tau_n) - q_j(l^{r-1}\tau_n)} \right)$$
behaves as
\[
\log \left( \frac{(ml)^{-\xi_j} - l^{-\xi_j}}{(m)^{-\xi_j} - 1} \right) = -\xi_j \log(l).
\]

The next theorem establishes the consistency and asymptotic normality of the estimator. For this purpose, I first extend the definition of the influence function in Theorem 2.3.1. In particular, for any positive constant \( k \), write
\[
\tilde{\phi}_{i,1,n}(k) := \frac{D_i}{P(X_i)} T_{i,1,n}(k) - \frac{E(T_{i,1,n}(k)|X_i)}{P(X_i)} (D_i - P(X_i))
\]
and
\[
\tilde{\phi}_{i,0,n}(k) := \frac{1 - D_i}{1 - P(X_i)} T_{i,0,n}(k) + \frac{E(T_{i,0,n}(k)|X_i)}{1 - P(X_i)} (D_i - P(X_i))
\]
where
\[
T_{i,1,n}(k) := k\tau_n - 1\{Y_{i,1} \leq q_{1}(k\tau_n)\}
\]
and
\[
T_{i,0,n}(k) := k\tau_n - 1\{Y_{i,0} \leq q_{0}(k\tau_n)\},\text{ respectively.}
\]

**Theorem 2.3.4.** Under the assumptions of Theorem 2.3.1, for \( j = 0, 1 \),

1. \( \hat{\xi}_j \xrightarrow{p} \xi_j \).
2. In addition, if
\[
\sqrt{\tau_n n} \left( \frac{-1}{\log(l)} \log \left( \frac{q_j(ml^r\tau_n) - q_j(l^r\tau_n)}{q_j(ml^{r-1}\tau_n) - q_j(l^{r-1}\tau_n)} \right) - \xi_j \right) \rightarrow 0
\]
as \( n \rightarrow \infty \) for all \( r = 1, 2, \ldots, R \), then, for \( b_r := \frac{(w_r - w_{r+1})l^r\xi_j(1-m^{-\xi_j})}{\log(l)\xi_j} \) and \( w_{R+1} = w_0 := 0 \), I have
\[
\sqrt{\tau_n n} (\hat{\xi}_j - \xi_j) = -\frac{1}{\sqrt{\tau_n n}} \sum_{i=1}^{n} \left( \sum_{r=0}^{R} b_r \left( \tilde{\phi}_{j,n,i}(ml^r) - \tilde{\phi}_{j,n,i}(l^r) \right) \right) + o_p(1).
\]
Denote \( \sigma^2_{j,n} := \text{Var} \left( \sum_{r=1}^{R} b_r \left( \hat{\phi}_{j,n,i}(m^r) - \tilde{\phi}_{j,n,i}(l^r) \right) \right) / \tau_n \), then

\[
\sqrt{\tau_n n \sigma_j^{-1}} (\hat{\xi}_j - \xi_j) \rightsquigarrow N(0,1).
\]

This theorem proves that the Pickands type estimator of the EV index is consistent. Under an additional assumption, its asymptotic normality holds too. The latter result can be used to test the type of tails of both \( Y_1 \) and \( Y_0 \).

2.4 Extreme quantile treatment effects

Section 2.4.1 establishes asymptotic theory for the \( \tau_n \)-th QTE when \( \tau_n \) is extreme. It serves as the foundation for the inference theory built in Sections 2.5.1 and 2.5.3. In addition, I will infer the 0-th QTE by a linear combination of extreme QTEs. Hence the asymptotic theory also contributes to the inference of 0-th QTE in Section 2.5.4. Section 2.4.2 verifies Assumption 8 which is a high-level assumption for the asymptotic theories of extreme QTE established in Section 2.4.1. Section 2.4.3 considers the asymptotic distribution of the extreme QTE estimator with a feasible normalizing factor. This permits inference through a resampling method proposed in Section 2.5.3.

2.4.1 The main result

Theorem 2.4.1 establishes an asymptotic theory for the \( \tau_n \)-th QTE when \( \tau_n \) is extreme. This gives the second main theoretical result of the paper: the asymptotic distribution of the estimator of extreme QTE is non-Gaussian.

First, assume the following,

Assumption 6. \( \tau_n \) is extreme; that is,

(1) \( \tau_n \to 0 \) as \( n \to \infty \),
(2) $\tau_n n \to k$ for some positive constant $k$ as $n \to \infty$.

Define the estimator $\hat{q}(\tau_n)$ of the $\tau_n$-th QTE $q(\tau_n)$ as:

$$\hat{q}(\tau_n) := \hat{q}_1(\tau_n) - \hat{q}_0(\tau_n) \quad (2.4.1)$$

where $\hat{q}_1(\tau_n)$ and $\hat{q}_0(\tau_n)$ are computed from (2.3.1) and (2.3.2), respectively.

In fact, I use the same objective functions as those used to compute the regular and intermediate QTE. On the practical side, this implies that researchers can compute them in a unified manner without pre-specification of which category the quantile index belongs to. On the theoretical side, I will show that the asymptotic behavior of $\hat{q}_j(\tau_n)$ is no longer normal comparing to the ones with intermediate and regular quantile indices. This is because the number of observations below $q_j(\tau_n)$ are of the same order of magnitude of $\tau_n n$, which does not diverge to infinity (Assumption 6). This also implies that I only need consistency of the propensity score estimator $\hat{P}_p x$.

**Assumption 7.** $\sup_{x \in \text{Supp}(X)} |\hat{P}(x) - P(x)| = o_p(1)$.

This assumption does not require the convergence rate for the nonparametric propensity score estimator is faster than $n^{1/4}$. The reason is similar: there are only a finite number of observations below the estimator of $\hat{q}_j(\tau_n)$, which are thus counted in the summation of (2.3.1) and (2.3.2). This prevents the accumulation of first order approximation error $\hat{P}(X_i) - P(X_i)$.

Next, I state a high-level assumption that determines the shape of the asymptotic distribution of the extreme QTE estimator.

**Assumption 8.** For $j = 0, 1$,

(1) $P(X \in \cdot | Y_j = y)$, the conditional distribution of $X$ given $Y_j = y$, weakly converges to the c.d.f. of a random variable $X_j$ as $y \to q_j(0)$. The c.d.f. of $X_j$ is denoted as
$P_j^+(X_j \in \cdot | Y_j = q_j(0))$.

(2) $P_j^+(X_j \in \cdot | Y_j = q_j(0))$ has finite mass points.

(3) Let $S$ be the discontinuity of $P(x)$. Then $P_j^+(X_j \in S | Y_j = q_j(0)) = 0$.

Assumption 8(1) is high-level. In Section 2.4.2, I provide primitive sufficient conditions for Assumption 8(1) to hold. Section A.1.1 contains some numerical illustrations of this condition. In general, $P_j^+(X_j \in \cdot | Y_j = q_j(0))$ depends on the structure of conditional boundary of $Y_j$ on $X$. The phenomenon that the asymptotic distribution depends on boundary conditions, is common in nonregular estimations. See, for example, Hirano and Porter (2003), Chernozhukov and Hong (2004), and Lee and Seo (2008). For Assumption 8(2), the number of mass points depends on the number of discrete minimizers of the conditional boundary of $Y_j$ given $X$ which is usually finite. Also, Assumption 8(2) does not rule out the case of $X$ being continuous, in which there is no mass point.

Theorem 2.4.1, the main theoretical result of this section, establishes the joint asymptotic distribution of $\hat{q}_j(\tau_n), j = 0, 1$ by showing that a normalized version of $\hat{q}_j(\tau_n), j = 0, 1$ weakly converges to the minimizer of an asymptotic objective function. I first state the normalized version of $\hat{q}_j(\tau_n), j = 0, 1$ below.

For $j = 0, 1$, the normalized versions of $\hat{q}_j(\tau_n)$ with or without centering are

$$\hat{Z}^c_{j,n}(k) := \alpha_{j,n}(\hat{q}_j(\tau_n) - q_j(\tau_n))$$

and

$$\hat{Z}_{j,n}(k) := \alpha_{j,n}(\hat{q}_j(\tau_n) - q_j^* - \beta_{j,n}),$$

respectively. Here, $q_j^*$ is an auxiliary constant so that $U_j = Y_j - q_j^*$ has lower endpoint 0 or $-\infty$. In particular, if $q_j(0) > -\infty$, then $q_j^* = q_j(0)$, otherwise, $q_j^*$ is arbitrary.
The normalizing constants \((\alpha_{j,n}, \beta_{j,n})\) for \(j = 0, 1\) are given by

- for type 1 tails \((\xi_j = 0): \alpha_{j,n} = 1/(a(F_{u_j}^{-1}(1/n))), \beta_{j,n} = F_{u_j}^{-1}(1/n),\)
- for type 2 tails \((\xi_j > 0): \alpha_{j,n} = -1/(F_{u_j}^{-1}(1/n)), \beta_{j,n} = 0,\)
- for type 3 tails \((\xi_j < 0): \alpha_{j,n} = 1/(F_{u_j}^{-1}(1/n)), \beta_{j,n} = 0,\)

in which \(F_{u_j}\) is the c.d.f. of \(U_j\).

Now I turn to the second part, the asymptotic objective function. The asymptotic objective function of the local parameter \(z\) takes the following form:

\[-kz + \sum_{i=1}^{\infty} W_j(D_{i,j}, P(X_{i,j}))l_\delta(\mathcal{J}_{i,j}, z), (2.4.2)\]

in which \(W_1(d, P) = \frac{d}{p}\) and \(W_0(d, P) = \frac{1-d}{1-p}\). To see the meaning of each term in (2.4.2), I denote, for \(j = 0, 1,\)

- for type 1 tails \((\xi_j = 0): h_j(l) = \exp(l)\), for \(l \in \mathbb{R}\), \(\eta_j(k) = \log(k),\)
- for type 2 tails \((\xi_j > 0): h_j(l) = (-l)^{-1/\xi_j}\), for \(l < 0\), \(\eta_j(k) = (-k)^{-\xi_j}\)
- for type 3 tails \((\xi_j < 0): h_j(l) = (l)^{-1/\xi_j}\), for \(l > 0\), \(\eta_j(k) = k^{-\xi_j}\).

Then \(\{\mathcal{E}_{i,j}, D_{i,j}, X_{i,j}\}\) is an i.i.d. sequence such that \(\{\mathcal{E}_{i,1}, D_{i,1}, X_{i,1}\} \perp \{\mathcal{E}_{i,0}, D_{i,0}, X_{i,0}\}\)
and for \(j = 0, 1, X_{i,j}\) is governed by the law \(P_j^+(X_j \in \cdot | Y_j = q_j(0)). D_{i,j}\) is Bernoulli distributed with success probability \(P(X_{i,j})\) conditional on \(X_{i,j}\) and \(\mathcal{E}_{i,j}\) is standard exponentially distributed independently of both \((X_{i,j}, D_{i,j})\). In addition, \(\mathcal{J}_{i,j} := h_j^{-1}(\sum_{i=1}^{n} \mathcal{E}_{i,j})\) and \(l_\delta(u, v) := \mathbb{1}\{u < v\}(v - u) - \mathbb{1}\{u \leq -\delta\}(-\delta - u)\) for an arbitrary \(\delta > 0\). The same function of \(l_\delta(u, v)\) is first used in Chernozhukov (2005a).

**Assumption 9.** For \(j = 0, 1\) and a generic fixed constant \(k > 0,\)

\[-kz + \sum_{i=1}^{\infty} W_j(D_{i,j}, P(X_{i,j}))l_\delta(\mathcal{J}_{i,j}, z)\]

has a unique minimizer almost surely.
Assumption 9 indicates that the asymptotic objective function has a unique minimizer which is necessary for applying the argmin theory. This type of assumption is common in nonregular estimation literature. See, for example, Chernozhukov and Fernández-Val (2011), Chernozhukov and Hong (2004), and Lee and Seo (2008). Lemma A.1.6 proves a sufficient condition for this assumption to hold. In general, the assumption holds when $X$ is absolutely continuous. If $X$ has a mass point at $x_0$, the sufficient condition requires that $kP(x_0)$ is not an integer, where $P(x)$ is the propensity score. Since integers are sparse on the real line, I consider this sufficient condition mild.

**Theorem 2.4.1.** If Assumptions 1, 20, 6–8 hold, there exist $\kappa_1$ and $\kappa_2$ such that $0 < \kappa_1 < \kappa_2 < \infty$ and $(\kappa_1, \kappa_2)$ satisfy Assumption 9, then $(\hat{Z}_{1,n}(k), \hat{Z}_{0,n}(k)) \rightsquigarrow (Z_{1,\infty}(k), Z_{0,\infty}(k))$ in $D^2([\kappa_1, \kappa_2])$, where

$$(Z_{1,\infty}(k), Z_{0,\infty}(k)) := \arg\min_{(z_1, z_0) \in \mathbb{R}^2} \sum_{j=0,1} \left[ -kz_j + \sum_{i=1}^{\infty} W_j(D_{i,j}, P(X_{i,j}))l_\delta(J_{i,j}, z_j) \right].$$

In addition, $(\hat{Z}_{1,n}^c(k), \hat{Z}_{0,n}^c(k)) \rightsquigarrow (Z_{1,\infty}^c(k), Z_{0,\infty}^c(k)) := (Z_{1,\infty}(k) - \eta_1(k), Z_{0,\infty}(k) - \eta_0(k))$ in $D^2([\kappa_1, \kappa_2])$.

The immediate corollary induced by the main theorem is the finite dimensional convergence. Due to the lack of continuity of the sample path of the asymptotic distribution, the projection mapping is only continuous when index $k$ is not at the discontinuity of the limiting process.

**Corollary 2.4.1.** If the assumptions in Theorem 2.4.1 hold and Assumption 9 is satisfied for $k \in \{k_i\}_{i=1}^L$, then

$$(\hat{Z}_{1,n}(k_i), \hat{Z}_{0,n}(k_i))_{i=1}^L \rightsquigarrow (Z_{1,\infty}(k_i), Z_{0,\infty}(k_i))_{i=1}^L := \arg\min_{(z_{1,i}, z_{0,i}) \in \mathbb{R}^2} \sum_{j=0,1} \sum_{l=1}^L \left[ -k_lz_{j,l} + \sum_{i=1}^{\infty} W_j(D_{i,j}, P(X_{i,j}))l_\delta(J_{i,j}, z_{j,l}) \right].$$
and

\[
(\hat{Z}_{1,n}^c(k_i), \hat{Z}_{0,n}^c(k_i))_{i=1}^L \rightsquigarrow (Z_{1,\infty}^c(k_i), Z_{0,\infty}^c(k_i))_{i=1}^L := (Z_{1,\infty}(k_i) - \eta_1(k_i), Z_{0,\infty}(k_i) - \eta_0(k_i))_{i=1}^L.
\]

First, this theorem, along with Theorem 2.3.1 (for the intermediate-order quantile) and Theorem 1 in Firpo (2007) (for the regular quantile), characterizes the evolution of the asymptotic distribution of the QTE estimator when the quantile index ranges from 0 to 1. Starting with the regular quantile, the asymptotic distribution is normal and estimating the unknown propensity score provides additional information. When the quantile index is intermediate, the shape of the asymptotic distribution remains normal, but the additional information from estimating the propensity score becomes asymptotically negligible if an additional envelope condition holds. When the quantile index moves even closer to the origin so that it is extreme, the shape of the asymptotic distribution becomes non-Gaussian, and the information from estimating the propensity score is asymptotically negligible. Figure 2.1 in Section 2.1 shows the evolution of the asymptotic distribution over quantile index \( \tau \).

Second, I do not impose any parametric restriction on the conditional quantile of \( Y_j \) given \( X \). This is in contrast to Chernozhukov (2005a), which considers linear extreme-order quantile regressions. The parameters considered in quantile regressions are conditional objects, while QTEs in this paper are unconditional objects. In order to deal with conditional quantiles, Chernozhukov (2005a) innovatively proposes to use the asymptotic independence between residuals and covariates \( X \) at tails in addition to linearity to regulate the conditional tail behavior. On the other hand, in this paper, I only need Assumption 8, which is a weaker assumption than linearity and asymptotic independence. Section 2.4.2 verifies Assumption 8 under three different conditional boundary conditions.

Third, I give another representation of the asymptotic objective function, which
will be used extensively in Section 2.4.2. In fact,

\[-kz + \sum_{i=1}^{\infty} W_j(D_{i,j}, P(X_{i,j})) I_d(J_{i,j}, z) = -kz + \int_{E_j} W_j(d, P(x)) I_d(u_j, z) dN_j(u_j, d, x),\]

where \(N_j(u_j, d, x)\) is a Poisson random measure on \(E_j\) with mean measure \(\mu_j(\operatorname{PRM}(\mu_j))\) and

for type 1 tails \((\xi_j = 0)\):
\[E_j = E^1 = [-\infty, +\infty) \times \{0, 1\} \times \operatorname{Supp}(\mathcal{X}),\]

for type 2 tails \((\xi_j > 0)\):
\[E_j = E^2 = [-\infty, 0) \times \{0, 1\} \times \operatorname{Supp}(\mathcal{X}),\]

for type 3 tails \((\xi_j < 0)\):
\[E_j = E^3 = [0, +\infty) \times \{0, 1\} \times \operatorname{Supp}(\mathcal{X}).\]

Let \(\mathcal{F}\) be a basis of relatively compact open sets of \(\mathbb{R}^r\) such that \(\mathcal{F}\) is closed under finite unions and intersections\(^3\) and for any \(F \in \mathcal{F}\), \(P_j^+(X_j \in \text{Bd}(F) | Y = q_j(0)) = 0\) in which \(\text{Bd}(F)\) is the boundary of the set \(F\). Then the mean measure \(\mu_j\), which uniquely determines the distribution of a Poisson random measure, is defined as

\[\mu_j((a,b) \times \{d\} \times F) := \int_F (dP(x) + (1 - d)(1 - P(x))) P_j^+(dx|Y_j = q_j(0))(h_j(b) - h_j(a)).\]

(2.4.3)

In Section 2.4.2, I establish the asymptotic distribution of \(\hat{q}_j(\tau_n)\) by deriving the close-form expressions the mean measure \(\mu_j\) under three scenarios.

Fourth, Theorem 2.4.1 has shown that \(\hat{q}_1(\tau_n)\) and \(\hat{q}_0(\tau_n)\) are asymptotically independent. This is because, by construction, \(\{J_{i,1}^{\prime}, X_{i,1}, D_{i,1}\}_{i \geq 1} \sqsupseteq \{J_{i,0}, X_{i,0}, D_{i,0}\}_{i \geq 1}\).

This implies that the joint asymptotic distribution of \((\hat{q}_0(\tau_n), \hat{q}_1(\tau_n))\) is fully characterized by the marginals. In Section A.1.1, I compute the marginal distribution of \(\hat{q}_1(\tau)\) under various boundary conditions.

Fifth, directly computing the critical value of the asymptotic distribution of \(\hat{q}(\tau_n)\) is infeasible. Note that the ultimate parameter of interest is \(q(\tau_n) := q_1(\tau_n) - q_0(\tau_n)\).

\(^3\) \(r\) is the dimension of \(X\).
Although the joint asymptotic distribution of \((\hat{q}_0(\tau_n), \hat{q}_1(\tau_n))\) has been established by Theorem 2.4.1, the convergences depend on the tails of \(Y_1\) and \(Y_0\) and are hard to estimate consistently. Furthermore, the asymptotic distributions of \(\hat{q}_0(\tau_n)\) and \(\hat{q}_1(\tau_n)\) are complicated and depend on unknown boundary conditions. In Section 2.5, I propose to use a \(b\) out of \(n\) bootstrap with or without replacement to construct a CI and to draw inferences.

Last, as pointed out in the first remark after Theorem 2.4.1, the shape of the asymptotic distribution changes as the quantile index moves from the intermediate region to the extreme region. The implication is that the finite sample distribution can be well approximated by the extreme-order quantile asymptotic distribution proposed in Theorem 2.4.1 only if \(k = \tau_n n\) is not large. In Section 4.5, I provide a detailed simulation to investigate this issue and find that an appropriate rule of thumb for referring to extreme-order quantile asymptotics to draw inference based on a \(b\) out of \(n\) bootstrap CI is

\[
\tau \leq \min\left(\frac{40}{n}, \frac{0.2b}{mn}\right)
\]

where \(m\) is the spacing parameter mentioned in Section 2.4.3.

### 2.4.2 Asymptotic distribution under various boundary conditions

This section verifies Assumption 8 under three different boundary conditions. I illustrate that the asymptotic distribution for the extreme QTE is nonregular and depends on the boundary condition in a complicated manner. More numerical illustrations are in Appendix A.1.1. Since the boundary condition is unknown and is usually hard to estimate, analytical inference is difficult. Instead, in Section 2.5, I will focus on resampling based inference, which does not require knowledge of the boundary.
The three different boundary conditions are based on the types of minimizers of the conditional boundary of $Y_j$ given $X$: finite minimizers, continuum minimizers, and mixture minimizers. I will restrict my attention to the marginal distribution of $\hat{q}_1(\tau_n)$ because of the asymptotic independence between $\hat{q}_1(\tau_n)$ and $\hat{q}_0(\tau_n)$.

**Finite minimizers**

I first consider the finite minimizers case. When the lower endpoint of $Y_1$ is bounded, I denote $\varpi(x)$ as $Y_1$’s conditional boundary given $X = x$. If $\varpi(x)$ is uniquely minimized at $x_0$, then as $Y_1 \to q_1(0)$, $X \to x_0$. So I expect $P^+_1(X_1 \in \cdot | Y_1 = q_1(0))$ to be $1\{x_0 \in \cdot \}$. This implies that the mean measure $\mu_1$ in the asymptotic distribution of $Z_{1,\infty}(k)$ defined in (2.4.3) takes the following form:

$$
\mu_1((a, b) \times \{d\} \times F) = (dP(x_0) + (1 - d)(1 - P(x_0)))(h_1(b) - h_1(a))1\{x_0 \in F\},
$$

for any $F \in \mathcal{F}$ in which

$$
\mathcal{F} := \text{a basis generated by all open sets in } \mathbb{R}^r \text{ containing } x_0 \text{ as an interior point.}
$$

Next, I will make the argument rigorous and generalize it to the scenario in which the conditional boundary achieves its minimum on finite points of $X$.

By the Skorohod representation in Lemma 7.11 of Van der Vaart (2000), there exists a measurable map $g$ on $\mathbb{R}^r \times [0, 1]$ and a uniformly $[0, 1]$ random variable $\varepsilon$ such that $Y_1 = g(X, \varepsilon)$, $X \perp \varepsilon$. On top of this, I assume:

**Assumption 10.** The measurable map $g$ is lower semi-continuous.

The conditional boundary obtains a finite set of minimizers; that is,

**Assumption 11.** $\varpi(x) > -\infty$ and is minimized at $S_0 = \{x_i\}_{i=1}^T$ for some positive integer $T < +\infty$. 

28
Now I characterize the weak limit $P_j^+ (X_j \in \cdot | Y_j = q_j(0))$ in Assumption 8 under Assumption 10 and 11. For each $y$, let $S_y$ be the support of random variable $\lambda(X, y)$ where $\lambda(x, y) := \Pr(g(x, \varepsilon) \leq y)$. For a fixed $y_0$, define $S_{y_0} := \cup_t S_{y_0,t}$ where $\{S_{y_0,t}\}$ is a partition of $S_{y_0}$ such that $x_t \in S_{y_0,t}$, $d(x_t', S_{y_0,t}) > 0$ for $t' \neq t$. For $y \leq y_0$, $S_{y,t} := S_{y_0,t} \cap S_y$ and $P_{y,t} := \frac{E1\{X \in S_{y,t}\} \frac{\lambda(X, y)}{\lambda(X, y')}}{E1\{X \in S_y\} \frac{\lambda(X, y)}{\lambda(X, y')}}$.

**Assumption 12.** $\lim_{y \to q_1(0)} P_{y,t}$ exists and is equal to $P_t$.

If Assumption 11 holds with $T = 1$, Assumption 12 holds with $P_1 = 1$ automatically.

Given Assumption 12, the asymptotic objective function is

$$-k z + \sum_{i=1}^\infty \frac{D_{i,1,f}}{P(X_{i,1,f})} l_d(J_{i,1,f}, z)$$

in which $\{E_{i,1,f}, D_{i,1,f}, X_{i,1,f}\}$ is a sequence of i.i.d. random vectors, $E_{i,1,f}$ is standard exponentially distributed, independent of $(X_{i,1,f}, D_{i,1,f})$, $J_{i,1,f} := h_1^{-1} \sum_{i=1}^\infty E_{i,1,f}$, $D_{i,1,f}$ is a Bernoulli distributed random variable with success probability $P(X_{i,1,f})$ conditional on $X_{i,1,f}$, and $X_{i,1,f}$ is supported by $S_0$ with corresponding point mass probabilities $\{P_t\}_{t=1}^T$. 29
Corollary 2.4.2. If Assumptions 1, 20, 6, 7, and 9-12 hold, then

\[ \hat{Z}_{1,n}(k) \rightsquigarrow Z_{1,\infty}(k) := \arg \min_{z \in \mathbb{R}} -kz + \sum_{i=1}^{\infty} \frac{D_{i,1,f}}{P(X_{i,1,f})} l_{i}(f_{i,1,f}, z). \]

Examples 1 and 2 in Section A.1.1 demonstrate the asymptotic distributions of this type.

Continuum minimizers

Next, I consider the conditional boundary in a case when it has continuum of minimizers; that is, a case in which it is flat over $X$. First, for $U_1 = Y_1 - q_1(0)$, I have

\[ P(X \in F | Y_1 = y) = \frac{\int_{F} f_{U_1}(y - q_1(0) | x) dF_X(x)}{\int_{F} f_{U_1}(y - q_1(0) | x) dF_X(x)}, \]

where $f_{U_1}$ is the conditional density of $U_1$. If $\varpi(x) > -\infty$ is flat, I can adapt the independence at infinity condition assumed in both Chernozhukov (2005a) and Chernozhukov and Fernández-Val (2011).

Assumption 13. $\varpi(x) \geq -\infty$ is flat, i.e. $\varpi(x) = q_1(0)$ and there exists a random variable $\varepsilon_1$ such that

1. For $u \to 0$, uniformly over $X$, $F_{U_1}(u | X) \sim F_{\varepsilon_1}(\frac{u}{\sigma_1(X)})$ and $f_{U_1}(u | X) \sim \frac{1}{\sigma_1(X)} f_{\varepsilon_1}(\frac{u}{\sigma_1(X)})$,
2. $\inf_x \sigma_1(x) > 0$,
3. $\xi_1$, the EV index of both $U_1$ and $\varepsilon_1$, is nonzero.
I allow the lower end-point to be $-\infty$. Assumption 13(1) means $U_1$ behaves as $\sigma_1(X)\varepsilon_1$ at its lower tail. We know that

$$f_{U_1}(u|X) \sim \frac{1}{\sigma_1(X)} f_{\varepsilon_1} \left( \frac{u}{\sigma_1(X)} \right)$$

holds point-wise by taking derivatives on both sides of $F_{U_1}(u|X) \sim F_{\varepsilon_1}(\frac{u}{\sigma_1(X)})$. See Resnick (1987) Proposition 0.7. Assumption 13(1) goes one-step further and requires that

$$f_{U_1}(u|X) \sim \frac{1}{\sigma_1(X)} f_{\varepsilon_1} \left( \frac{u}{\sigma_1(X)} \right)$$

holds uniformly. The uniformity is not strong, given that $\text{Supp}(X)$ is compact. It can be relaxed to hold point-wisely with an envelope condition as illustrated in D’Haultfoeuille et al. (2015). Based on Assumption 13,

$$f_{U_1}(y - q_1(0)|X) \sim \frac{1}{\sigma_1(X)} f_{\varepsilon_1} \left( \frac{y - q_1(0)}{\sigma_1(X)} \right) \sim \sigma_1(X)^{1/\xi_1} f_{\varepsilon_1}(y - q_1(0))$$

uniformly over $X$.

Under the conditional independence at the tail, as $y \to q_1(0)$, I have

$$P(X \in F|Y_1 = y) \to \frac{\int_F \sigma_1(x)^{1/\xi_1} dF_X(x)}{\int_{\text{Supp}(X)} \sigma_1(x)^{1/\xi_1} dF_X(x)}.$$
Then, the asymptotic objective function is

$$-kz + \sum_{i=1}^{\infty} \frac{D_{i,1,c}}{P(X_{i,1,c})} l_b(J_{i,1,c}, z)$$

in which \{X_{i,1,c}, D_{i,1,c}, E_{i,1,c}\} is i.i.d. sequence of random vectors, \(X_{i,1,c}\) is generated from the density

$$\frac{\sigma_1(x)^{1/\xi} dF_X(x)}{\int_{\text{supp}(X)} \sigma_1(x)^{1/\xi} dF_X(x)},$$

\(D_{i,1,c}\) is Bernoulli distributed with success probability \(P(X_{i,1,c})\) conditional on \(X_{i,1,c}\), \(E_{i,1,c}\) is a standard exponentially distributed random variable that is independent of \(X_{i,1,c}\) and \(D_{i,1,c}\), and \(J_{i,1,c} := h_1^{-1}(\sum_{l=1}^{i} E_{i,1,c}).

**Corollary 2.4.3.** If Assumptions 1, 20, 6, 7, 9 and 13 hold,

$$\hat{Z}_{1,n}(k) \xrightarrow{\text{a.s.}} Z_{1,\infty}(k) := \arg\min_{z \in \mathbb{R}} -kz + \sum_{i=1}^{\infty} \frac{D_{i,1,c}}{P(X_{i,1,c})} l_b(J_{i,1,c}, z).$$

Example 3 in Section A.1.1 illustrates this type of asymptotic distribution.

**Mixture Minimizers**

Last, I combine the two types of boundary structures and consider the case in which the minimizers of the conditional boundary is a mixture of discrete points and continuum intervals. For two positive integers \(T^c\) and \(R^d\), let \(\varpi(x) > -\infty\) achieve its minimum on

$$x \in \{x_1, \cdots, x_{R^d}\} \cup (\cup_{t=1}^{T^c} S_{0,t}).$$

For each \(y\), let \(S_y\) be the support of random variable \(\lambda(X, y)\) where

$$\lambda(x, y) := \Pr(g(x, \varepsilon) \leq y).$$
And for fixed $y_0$, let
\[
\{ \{ S^d_{y_0,t} \} \}_{r=1}^{R^d}, \{ S^c_{y_0,t} \} \}_{t=1}^{T^c}
\]
be a partition of $S_{y_0}$ such that
\[
x_r \in S^d_{y_0,r}, \quad S^c_{0,t} \subset S^c_{y,t}, \quad d(x_r, S^d_{y_0,r}) > 0
\]
for all $r \neq 'r'$, $d(S^c_{y_0,t}, S^d_{y_0,r}) > 0$ for all $t$ and $r$, and $d(S^c_{y_0,t}, S^c_{y_0,t'}) > 0$ for all $t \neq t'$. Then let
\[
S^d_{y,r} := S^d_{y_0,r} \cap S_y, \quad P^d_{y,r} := \frac{E \{ \xi \in S^d_{y,r} \} \frac{\partial \lambda(x,y)}{\partial y}}{E \{ \xi \in S_y \} \frac{\partial \lambda(x,y)}{\partial y}}
\]
for $r = 1, 2, \cdots, R^d$ and
\[
S^c_{y,t} := S^c_{y_0,t} \cap S_y, \quad P^c_{y,t} := \frac{E \{ \xi \in S^c_{y,t} \} \frac{\partial \lambda(x,y)}{\partial y}}{E \{ \xi \in S_y \} \frac{\partial \lambda(x,y)}{\partial y}}
\]
for $t = 1, 2, \cdots, T^c$.

**Assumption 14.**

(1) For $d(\cdot, \cdot)$, the Euclidean distance between sets or between points and sets,
\[
\min_{r \neq 'r'} d(x_r, x_{r'}) \wedge \min_{t \neq 't'} d(S_{0,t}, S_{0,p}) \wedge \min_{r \in R^d, t \in T^d} d(x_r, S_{0,t}) > \delta_0
\]
for some positive $\delta_0$.

(2) As $y \to q_1(0)$, $P^d_{y,r} \to P^d_r$ for $r = 1, 2, \cdots, R^d$ and $P^c_{y,t} \to P^c_t$ for $t = 1, 2, \cdots, T^c$.

(3) Let $S^\delta$ denote the $\delta$-enlargement set $\{ x | d(x, S) \leq \delta \}$; there then exists a positive constant $\delta$ such that for each $t = 1, 2, \cdots, T^c$, on $(S_{0,t})^\delta$, there exist $\xi_t$ with EV index $\xi_t < 0$ and $\sigma_t$ such that the following relation holds:
\[
f_{\xi_t}(y - q_1(0) | X = x) \sim \frac{1}{\sigma_t(x)} f_{\xi_t}(y - q_1(0)) \sim \sigma_t(x)^{-1/\xi_t} f_{\xi_t}(y - q_1(0))
\]
uniformly in $x \in (S_{0,t})^\delta$.

(4) $\min_{t \in T^c} \inf_x \sigma_t(x) > 0.$
Next I define the asymptotic objective function for the mixture boundary case:

$$-kz + \sum_{i=1}^{\infty} \frac{D_{i,1,m}}{P(\mathcal{X}_{i,1,m})} l_\delta(J_{i,1,m}, z),$$

in which \(\mathcal{E}_{i,1,m}, \mathcal{D}_{i,1,m}, \mathcal{X}_{i,1,m}\) is an i.i.d. sequence of random vectors, \(\mathcal{E}_{i,1,m}\) is standard exponentially distributed, independent of both \(\mathcal{X}_{i,1,m}\) and \(\mathcal{D}_{i,1,m}, J_{i,1,m} := h_1^{-1}(\sum_{t=1}^{i} \mathcal{E}_{t,1,m})\), \(\mathcal{D}_{i,1,m}\) is Bernoulli distributed with success probability \(P(\mathcal{X}_{i,1,m})\) conditional on \(\mathcal{X}_{i,1,m}\), \(\mathcal{X}_{i,1,m}\) is supported on \(\{x_1, \ldots, x_{R^d}\} \cup (\cup_{t=1}^{T} S_0, t)\), with its distribution being that, for any Borel set \(B\),

$$P(\mathcal{X}_{i,1,m} \in B) = \sum_{r=1}^{R^d} 1\{x_r \in B\} P_r^d + \sum_{t=1}^{T} \int_{S_0,t \cap B} \sigma_t(x)^{1/\xi_t} dF_X(x) / \sigma_t(x)^{1/\xi_t} dF_X(x).$$

**Corollary 2.4.4.** If Assumptions 1, 20, 6, 7, 9, 10 and 14 hold, then

$$\hat{Z}_{1,n}(k) \xrightarrow{\ast} Z_{1,\infty}(k) := \arg\min_{z \in \mathbb{R}} -kz + \sum_{i=1}^{\infty} \frac{D_{i,1,m}}{P(\mathcal{X}_{i,1,m})} l_\delta(J_{i,1,m}, z).$$

Example 4 in Section A.1.1 describes this type of asymptotic distribution.
2.4.3 Feasible normalizing factor

Sections 2.4.1 and 2.4.2 characterize the asymptotic distribution for the extreme QTE under high-level and primitive conditions. This section considers the next missing piece needed for the resampling inference method: the feasible normalizing factor. I propose a feasible normalizing factor that is not a consistent estimator but has the same order of magnitude as the infeasible one and establish the corresponding asymptotic theory.

The normalizing factor for the $\tau_n$-th QTE estimator when $\tau_n$ is extreme has not been obvious. First, the estimator of $\tau_n$-th QTE is $\hat{q}(\tau_n) := \hat{q}_1(\tau_n) - \hat{q}_0(\tau_n)$. Due to the different tail behaviors, the normalizing factors for $\hat{q}_1(\tau_n)$ and $\hat{q}_0(\tau_n)$ are not necessarily the same. In addition, by Theorem 2.4.1, the normalizing factors for $\hat{q}_1(\tau_n)$ and $\hat{q}_0(\tau_n)$ are first-order statistics that are unknown and hard to estimate.

I propose the following feasible normalizing factor:

$$\hat{\alpha}_n := \frac{\sqrt{\tau_{n,v} n}}{\max\left\{\hat{q}_1(m\tau_{n,v}) - \hat{q}_1(\tau_{n,v}), \hat{q}_0(m\tau_{n,v}) - \hat{q}_0(\tau_{n,v})\right\}}, \quad (2.4.4)$$

where $m$ is a spacing parameter. The feasible normalizing factor uses the smaller of the two factors for $\hat{q}_1(\tau_n)$ and $\hat{q}_0(\tau_n)$. In addition, the proposed factor has the same order but is not a consistent estimator of the infeasible order statistic. This is possible by the following assumption:

Assumption 15.

(1) $\tau_{n,v} n \to k_v$.

(2) $k_v$ satisfies the condition in Lemma A.1.7 as well as Assumption 9.

(3) Both $Y_1$ and $Y_0$ have type 2 or 3 tails.

This assumption is valid in many economic applications. First, type 2 or 3 tails are also called Pareto-type tails, which are prevalent in economic data such as wealth
and incomes, as argued in Section 2.2 of Chernozhukov and Fernández-Val (2011). Theoretically, it implies that c.d.f. decays polynomially. Second, the assumption holds if and only if the EV index is non-zero, which is testable based on Theorem 2.3.4. In practice, it implies that the c.d.f. of the two potential outcomes decay polynomially as $\tau \to 0$. Last, it implies that the feasible and infeasible normalizing factors are of the same order of magnitude. To see this, with $n \to \infty$, I have

$$
\frac{q_j\left(\frac{1}{n}\right)}{q_j(m\tau_{n,p}) - q_j(\tau_{n,p})} \sim \frac{k^n_{j\xi_j}}{m^n - 1}.
$$

Theoretically, the choice of $\tau_{n,p}$ in $\hat{\alpha}_n$ does not impact the asymptotic validity of the normalizing factor. However, in finite sample, this choice involves a trade-off between bias and variance. If $n\tau_{n,p}$ is small, there are fewer observations used for estimating $\hat{q}_j(\tau_{n,p})$, which produces a large variance. On the other hand, if $n\tau_{n,p}$ is large, it can introduce bias. First, in this case, the estimation error of the propensity score will accumulate and contaminate the CI. In addition, since I use a $b$ out of $n$ bootstrap method with subsample size $b$ to construct the CI, if $mn\tau_{n,p}/b$ is large, then this quantile index cannot be interpreted as extreme-order. Both implies that the EV asymptotic theory is not suitable. To address all the issues aforementioned, The rule of thumb I use to choose the index $\tau_{n,p}$ is $\tau_{n,p} = \min\left(\frac{C_1}{n}, \frac{C_2b}{mn}\right)$. The simulation study in Section 4.5 shows that this rule with $(C_1, C_2) = (10, 0.1)$ performs well in finite samples.

Similar to Assumption 5, I have to bridge the two normalizing factors.

**Assumption 16.** \(\frac{q_1(m\tau_n) - q_1(\tau_n)}{\hat{q}_0(m\tau_n) - \hat{q}_0(\tau_n)} \to \rho(m, k) \text{ for } \tau_n \to k \text{ uniformly in } k \in [\kappa_1, \kappa_2] \text{ where } \rho(m, k) \in [0, \infty].\)

This assumption bridges the convergence rates for $\hat{q}_1(\tau_n)$ and $\hat{q}_0(\tau_n)$ in the same spirit of Assumption 5. Since $\rho(m, k)$ is allowed to take $0$ and $\infty$, the assumption incorporates the case when one convergence rate dominates another.
The next theorem characterizes the weak convergence of the extreme QTE estimator with the feasible normalizing factor.

**Theorem 2.4.2.** The assumptions in Theorem 2.4.1 and Assumptions 15 and 16 hold. Note that

\[ \hat{\rho}(m, k) := \rho(m, k) k^{\xi_1 - \xi_0} \frac{m^{-\xi_0} - 1}{m^{-\xi_1} - 1} \quad \text{and} \quad \hat{Z}_n^c(k) := \hat{\alpha}_n(q(\tau_n) - q(\tau_n)) \]

for any \( \tau_n \to k \). Then for \( k \) fixed,

\[ \hat{Z}_n^c(k) \Rightarrow Z_\infty^c(k) \quad \text{in} \quad D[\kappa_1, \kappa_2], \]

in which

\[ Z_\infty^c(k) := \frac{\sqrt{k_\|}(Z_{1,\infty}^c(k) - \hat{\rho}(m, k) Z_{0,\infty}^c(k))}{\max \left\{ Z_{1,\infty}(mk) - Z_{1,\infty}^c(k), \hat{\rho}(m, k)(Z_{0,\infty}(mk) - Z_{0,\infty}(k)) \right\}}. \]

An immediate corollary from the above theorem is the weak convergence of a linear combination of \( \hat{Z}_n^c(k) \)'s. In Section 2.5.4, I use the linear combination of extreme QTE estimators to construct a point estimator and a CI for the 0-th QTE. Proposition 2.4.5 establishes the theoretical foundation for this construction. The key here is to choose a proper set of weights \( \{ \hat{r}_l \}_{l=1}^L \), which is explained in Section 2.5.4.

**Assumption 17.** Let \( \{ \hat{r}_l \}_{l=1}^L \) be a set of weights that can be random, and

1. \( \sum_{l=1}^L \hat{r}_l = 1 \),
2. \( \hat{r}_l \xrightarrow{p} r_l \) for all \( l = 1, \cdots, L \) and \( \{ r_l \}_{l=1}^L \) a set of constant real numbers.
3. \( \tau_{n,l,n} \to k_l \) where \( \{ k_l \}_{l=1}^L \) satisfies Assumption 9.
Corollary 2.4.5. The assumptions in Theorem 2.4.2 and Assumption 17 hold. Then
\[ \hat{\alpha}_n \left( \sum_{l=1}^{L} \hat{r}_l \hat{q}(\tau_{n,l}) - \sum_{l=1}^{L} r_l q(\tau_{n,l}) \right) \rightsquigarrow \sum_{l=1}^{L} r_l Z_{\tau_n}^c(l_i). \]

2.5 Inference

This section establishes inference theory for extreme QTE estimators that I apply in Section 2.7. Section 2.5.1 shows that the conventional bootstrap CI is inconsistent. Section 2.5.3 establishes a new uniformly consistent CI over a range of quantile indices. Last, Section 2.5.4 proposes to infer the 0-th QTE by combining a set of extreme QTE estimators with carefully chosen weights.

2.5.1 Inconsistency of the standard bootstrap inference method

Based on the asymptotic theory established in Section 2.4, this section shows that the conventional bootstrap CI, that is, the \( n \) out of \( n \) bootstrap CI (NN-CI) is inconsistent. This inconsistency explains the poor performance of the conventional bootstrap CI documented in Section 4.5.

Theorem 2.5.1 shows that the asymptotic distribution of the conventional bootstrap estimator is different from the one of the original estimator established in Theorem 2.4.1. I first define the bootstrap estimator with proper normalizations:

\[
(\hat{Z}_{1,n}^*(k), \hat{Z}_{0,n}^*(k)) := \arg\min_{(z_1, z_2) \in \mathbb{R}^2} \sum_{j=0,1} \left\{ -\sum_{i=1}^{n} \left( \sum_{l=1}^{n} 1\{ I_l = i \} \right) W_j(D_i, \hat{P}(X_i)) \alpha_{j,n}(U_{i,j} - q_j(0)), z_j \right\}
\]

\[
+ \sum_{i=1}^{n} \left( \sum_{l=1}^{n} 1\{ I_l = i \} \right) W_j(D_i, \hat{P}(X_i)) l_\delta(\alpha_{j,n}(U_{i,j} - q_j(0)), z_j)
\]

in which \( \hat{Z}_{j,n}^*(k) := \alpha_{j,n}(\hat{q}_{j,n}(\tau_n) - q_j(0)) \) for \( \tau_n \rightarrow k \). \( \hat{q}_{j,n}(\tau_n) \) is the point estimator computed from (2.3.1) and (2.3.2) using the bootstrap sample. Similarly,
\[ \hat{Z}_{j,n}^\alpha(k) := \alpha_{j,n}(\hat{q}_{j,n}^\tau(\tau_n) - q_j(\tau_n)). \] Here, \((I_{n,1}, I_{n,2}, \ldots, I_{n,n})\) is a multinomial vector with parameter \(n\) and probabilities \(\frac{1}{n}, \ldots, \frac{1}{n}\). The data is denoted as \(\Phi_n\) and \((I_{n,1}, I_{n,2}, \ldots, I_{n,n}) \perp \Phi_n\).

**Theorem 2.5.1.** The Assumptions in Theorem 2.4.1 hold. Then

\[ (\hat{Z}_{1,n}^*(k), \hat{Z}_{0,n}^*(k)) \rightsquigarrow (Z_{1,\infty}^*(k), Z_{0,\infty}^*(k)), \]

in which

\[ (Z_{1,\infty}^*(k), Z_{0,\infty}^*(k)) := \arg \min_{(z_1, z_0) \in \mathbb{R}^2} \sum_{j=0}^{\infty} \left[ -k z_j + \sum_{i=1}^{\infty} \Gamma_{i,j} W_j(D_{i,j}, P(X_{i,j})) l_\delta(J_{i,j}, z_j) \right] \]

and

\[ (\hat{Z}_{1,n}^{\alpha^*}(k), \hat{Z}_{0,n}^{\alpha^*}(k)) \rightsquigarrow (Z_{1,\infty}^{\alpha^*}(k), Z_{0,\infty}^{\alpha^*}(k)) := (Z_{1,\infty}^*(k) - \eta_1(k), Z_{0,\infty}^*(k) - \eta_0(k)). \]

Here, \(\{J_{i,j}, D_{i,j}, X_{i,j}\}_{i \geq 1, j = 0, 1}\) are the same as in Theorem 2.4.1 and \(\{\Gamma_{i,j}\}_{i \geq 1}\) is a sequence of i.i.d. Poisson random variables with unit mean such that

\[ \{\Gamma_{i,j}\}_{i \geq 1, j = 0, 1} \perp \{J_{i,j}, D_{i,j}, X_{i,j}\}_{i \geq 1, j = 0, 1} \]

and \(\Gamma_{i,1} \perp \Gamma_{i,0}\).

The asymptotic distribution of the bootstrap estimator of extreme QTE is different from the original estimator. Compared with the limiting process in Theorem 2.4.1, there is an additional Poisson random variable term. Since the asymptotic objective function is not quadratic, \(Z_{j,\infty}^\alpha, j = 0, 1\) are not linear in \(\Gamma_{i,j}\) which causes the invalidity of the bootstrap inference. Furthermore, due to the lack of linear expansion of the estimator, \(\hat{Z}_{j,n}^*(k) - \hat{Z}_{j,n}(k)\) does not share the same limiting distribution with \(\hat{Z}_{j,n}(k)\).

The intuition behind the invalidity of standard bootstrap here is similar to the case of order statistics. When there are no missing counterfactuals, that is, the data
are fully missing at random, the extreme-order quantile estimator considered in this paper degenerates to an order statistic. However, Bickel and Freedman (1981) has already shown that the standard $n$ out of $n$ bootstrap inference is not consistent for order statistics.

### 2.5.2 Consistency of the $b$ out of $n$ bootstrap inference

We have just seen that the conventional bootstrap CI is inconsistent. In this section, I establish the uniform consistency of a $b$ out of $n$ bootstrap CI (BN-CI) both with and without replacement in which $b$ is the subsample size with $b \to \infty$, $\frac{b}{n} \to 0$. This is the third main theoretical result of the paper. It allows empirical researchers to do uniformly consistent inferences over an range of quantile indices. Section 4.5 confirms the consistency of BN-CI as well as the inconsistency of NN-CI through an extensive numerical study.

Let the quantile index for the subsample be $\tau_b$. The key insight for the $b$ out of $n$ bootstrap inference both with and without replacement is to align $\tau_b$ with $\tau_n$. Theorem 2.4.2 shows that the asymptotic distribution of the $\tau_n$-th QTE is indexed by $k$. Letting $\tau_b = \tau_n n = k$ ensures that the subsample estimator can mimic the same asymptotic distribution of the full sample estimator.

I consider the $b$ out of $n$ bootstrap inference for extreme QTEs both with and without replacement. Not allowing for replacement (subsampling), Bertail et al. (2004) studied the validity of inference for extreme-order statistics without covariates. Chernozhukov and Fernández-Val (2011) considered a similar inference procedure in linear extreme-order quantile regressions. Allowing for replacement, Bickel and Sakov (2008) considered the $b$ out of $n$ bootstrap inference in extreme-order statistics without covariates. Theorem 2.5.2 proves the consistency of $b$ out of $n$ bootstrap inference both with and without replacement for the extreme QTE.\footnote{I suggest using the $b$ out of $n$ bootstrap with replacement because it performs better in simulation.}
Before stating the main theorem of this section, I introduce the resampling version of the feasible normalizing factor

\[
\hat{\alpha}_b^* := \frac{\sqrt{\tau_{b,b'} b}}{\max\left\{\hat{q}_1^b(m\tau_{b,b'}) - \hat{q}_0^b(m\tau_{b,b'}), \hat{q}_0^b(m\tau_{b,b'}) - \hat{q}_0^b(\tau_{b,b'})\right\}}
\]

where \(\tau_{b,b'} = \tau_{n,l} n, \tau_{n,l}\) satisfies Assumption 15, and the normalized estimator

\[
\hat{Z}_{c,n}^c(k) := \hat{\alpha}_b^* (\hat{q}_0^b(\tau_b) - \hat{q}(\tau_b)).
\]

In the above two equations, \(\hat{q}_0^b := \hat{q}_1^b(\tau_b) - \hat{q}_0^b(\tau_b)\) where \(\hat{q}_j^b(\tau_b)\) is computed by (2.3.1) and (2.3.2) with \(\tau_n\) replaced by \(\tau_b\) and using only the data from the subsample, which is generated either with or without replacement. Without the star symbol, \(\hat{q}(\tau_b) := \hat{q}_1(\tau_b) - \hat{q}_0(\tau_b)\) where \(\hat{q}_j(\tau_b)\) is computed by (2.3.1) and (2.3.2) with \(\tau_n\) replaced by \(\tau_b\) and using the full sample.

**Theorem 2.5.2.** If the assumptions in Theorem 2.4.2 hold and as \(n \to \infty, \frac{b}{n} \to 0, b \to \infty\) at a polynomial rate in \(n\), then \(\hat{Z}_{c,n}^c(k) \rightsquigarrow Z_{\infty}^c(k)\) in \(D([\kappa_1, \kappa_2])\).

Theorem 2.5.2 builds the theoretical foundation for constructing the uniform confidence band for the extreme QTE over \(k \in [\kappa_1, \kappa_2]\), in which \(\kappa_1, \kappa_2\) are not at the discontinuity of the limiting process with probability 1. To construct a uniformly consistent confidence band, I next want to studentize the process \(\hat{Z}_{n}^c(k)\). When the limiting process is Gaussian, it is common to studentize the process first and then to approximate the studentized limit. Here, I consider the same studentization in the non-Gaussian case. Let \(S_n(k)\) and \(\sigma(k)\) be the feasible and infeasible studentizing factors.

**Assumption 18.** For a (random) scale function \(S_n(k)\), there exists \(\sigma(k) > 0\), a deterministic function of \(k\), such that

\[
\sup_{k \in [\kappa_1, \kappa_2]} \left| \frac{S_n(k)}{\sigma(k)} - 1 \right| = o_p(1).
\]
In addition, with probability approaching one, \( \sigma(k) \), \( S_n(k) \) are both continuous in \( k \) and uniformly bounded and bounded away from zero over \( k \in [\kappa_1, \kappa_2] \).

\( S_n(k) \) can be \( S_n(k) := 1 \) or \( S_n(k) := k^{-\xi_1} + k^{-\xi_2} \) with corresponding \( \sigma(k) := 1 \) or \( \sigma(k) := k^{-\xi_1} + k^{-\xi_2} \), respectively. In the later case, \( \xi_j \), \( j = 0, 1 \) are unknown. So I replace them by their consistent estimators \( \hat{\xi}_j \), \( j = 0, 1 \). The choice of studentizing factors will not affect the size of the uniform confidence band, but will rather affect its power. Unlike the Gaussian limit in which using \( \sigma(k) \) as the point-wise standard deviation is natural, the best choice for the studentizing factor in this non-Gaussian case is still an open question and should be the focus of future research.

**Corollary 2.5.1.** Let \( \hat{C}_{1-a} \) denote the \((1-a)\)-th quantile of \( \max_{k \in [\kappa_1, \kappa_2]} |\hat{Z}^*_n(k)/S_n(k)| \).

If the assumptions in Theorem 2.5.2 and Lemma A.1.7 as well as Assumption 18 hold, then

\[
P \left( q \left( \frac{k}{n} \right) \in \left[ \hat{q} \left( \frac{k}{n} \right) - S_n(k) \hat{C}_{1-a}/\hat{\alpha}_n, \hat{q} \left( \frac{k}{n} \right) + S_n(k) \hat{C}_{1-a}/\hat{\alpha}_n \right] : k \in [\kappa_1, \kappa_2] \right) \to 1-a.
\]

Let \( \{k_i\}_{i=1}^L \) is a fine grid, \( \tau_{n,i} = \frac{k_i}{n} \), \( \tau_{b,l} = \frac{k_i}{b} \), and the number of subsamples be \( B_n \), which is as large as computationally possible. Researchers can compute the uniform confidence band based on the following procedure.

1. Compute \( \hat{q}(\tau_{n,i}) \) and \( \hat{q}(\tau_{b,l}) \) as in (2.4.1). Compute \( \hat{\alpha}_n \), \( S_n(k) \), and the propensity score using the full sample.

2. For the \( i \)-th subsample, compute \( \hat{q}_{b,1}^*(\tau_{b,l}) \) for \( l = 1, \cdots, L \) as in (2.4.1). Denote

\[
\hat{\alpha}_{i,b}^* := \frac{\sqrt{\tau_{b,l}^b}}{\max\left\{ \hat{q}_{i,1}^*(m\tau_{b,l}) - \hat{q}_{i,1,b}^*(\tau_{b,l}), \hat{q}_{i,0,b}^*(m\tau_{b,l}) - \hat{q}_{i,0,b}^*(\tau_{b,l}) \right\}}
\]
where for $j = 0, 1$, $\hat{q}^{*}_{i,j,b}(\tau)$ is computed as in (2.4.1) for each subsample with $\tau = m\tau_{b,l}$ or $\tau_{b,l}$. Denote $\hat{V}^{*}_{i,b} := \max_{i=1,\ldots,L} \hat{\alpha}^{*}_{i,b} \left| \left( \hat{q}^{*}_{i,b}(\tau_{b,l}) - \hat{q}(\tau_{b,l}) \right) / S_n(k) \right|$.  

3. Compute $\hat{C}_{1-a}$ as the $(1-a)$-th quantile of the $\{\hat{V}^{*}_{i,b}\}_{i=1}^{B}$.

Next I consider the $b$ out of $n$ inference for a linear combination of extreme QTEs. By carefully choosing the weights, in Section 2.5.4, I show that the linear combination of extreme QTE estimators can be utilized to infer the 0-th QTE.

Let $C_a$ be the $a$-th quantile of $\sum_{l=1}^{L} \gamma_r Z_{\infty}^c(k_l)$ and $\hat{C}_a$ be the $a$-th quantile of

$$
\hat{\alpha}_b \left( \sum_{l=1}^{L} \hat{\gamma}_l \hat{q}^{*}_b(\tau_{b,l}) - \sum_{l=1}^{L} \hat{\gamma}_l \hat{q}(\tau_{b,l}) \right).
$$

Given that $\sum_{l=1}^{L} \gamma_r Z_{\infty}^c(k_l)$ is continuous, Proposition 2.5.1 shows that $\hat{C}_a$ is a consistent estimator of $C_a$. Denote

$$
\sum_{l=1}^{L} \hat{r}_l \hat{q}(\tau_{n,l}) - \hat{C}_{0.5}/\hat{\alpha}_n \quad \text{and} \quad \left[ \sum_{l=1}^{L} \hat{r}_l \hat{q}(\tau_{n,l}) - \hat{C}_{1-a/2}/\hat{\alpha}_n, \sum_{l=1}^{L} \hat{r}_l \hat{q}(\tau_{n,l}) - \hat{C}_{a/2}/\hat{\alpha}_n \right]
$$

the median-unbiased estimator and a $(1-a) \times 100\%$ CI for $\hat{q}(\tau)$, respectively.

**Proposition 2.5.1.** Under the assumptions in Theorem 2.5.2 and Assumption 17, I have

$$
\hat{\alpha}_b \left( \sum_{l=1}^{L} \hat{\gamma}_l \hat{q}^{*}_b(\tau_{b,l}) - \sum_{l=1}^{L} \hat{\gamma}_l \hat{q}(\tau_{b,l}) \right) \overset{\text{Law}}{\rightarrow} \sum_{l=1}^{L} \gamma_r Z_{\infty}^c(k_l), \quad (2.5.1)
$$

$$
\lim_{n \to \infty} P \left( \sum_{l=1}^{L} \hat{r}_l \hat{q}(\tau_{n,l}) - \hat{C}_{0.5}/\hat{\alpha}_n \leq \sum_{l=1}^{L} r_l q(\tau_{n,l}) \right) = 0.5, \quad (2.5.2)
$$

and

$$
\lim_{n \to \infty} P \left( \sum_{l=1}^{L} \hat{r}_l \hat{q}(\tau_{n,l}) - \hat{C}_{1-a/2}/\hat{\alpha}_n \leq \sum_{l=1}^{L} r_l q(\tau_{n,l}) \leq \sum_{l=1}^{L} \hat{r}_l \hat{q}(\tau_{n,l}) - \hat{C}_{a/2}/\hat{\alpha}_n \right) = 1 - a. \quad (2.5.3)
$$

5 This is shown in Lemma A.1.7 in the appendix.

43
Here, (2.5.1) shows the weak convergence of the linear combination of extreme QTE estimators. (2.5.2) shows the median-unbiased estimator is asymptotically median-unbiased and (2.5.3) implies that the CI asymptotically controls size.

Let $B_n$ denote the number of subsamples. I use the following steps to compute $\hat{C}_a$.

1. Compute $\{\hat{r}_l\}_{l=1}^L, \hat{q}(\tau_{b,l}), \hat{q}(\tau_{n,l})$, and the propensity score estimator $\hat{P}(x)$ using the full sample.

2. For the $i$-th subsample, compute $\hat{q}_{i,b}^*(\tau_{b,l})$ for $l = 1, \cdots, L$ as in (2.4.1). Denote

$$\hat{\alpha}_{i,b}^* := \frac{\sqrt{\tau_{b,l} b}}{\max \left\{ \hat{q}_{i,1,b}^*(m\tau_{b,l}) - \hat{q}_{i,1,b}^*(\tau_{b,l}); \hat{q}_{i,0,b}^*(m\tau_{b,l}) - \hat{q}_{i,0,b}^*(\tau_{b,l}) \right\}}$$

where for $j = 0, 1, \hat{q}_{i,j,b}^*(\tau_b)$ is computed as in (2.4.1) for each subsample. Denote

$$\hat{V}_{i,b}^* := \hat{\alpha}_{i,b}^* \sum_{l=1}^L \hat{r}_l (\hat{q}_{i,b}^*(\tau_{b,l}) - \hat{q}(\tau_{b,l}))$$

3. Compute $\hat{C}_{1-a}$ as the $(1-a)$-th quantile of the $\{\hat{V}_{i,b}^*\}_{i=1}^{B_n}$.

When $L = 1$, I can use this procedure to construct the CI for $\hat{q}(\tau_n) := \hat{q}_1(\tau_n) - \hat{q}_0(\tau_n)$, the estimator of the $\tau_n$-th QTE. The finite sample performance of the CI is examined in the second part of Section 4.5. When $L \geq 2$, I can consider a finite combination of extreme QTE estimators with generic random weights. If the weights are chosen based on the EV indices for $Y_1$ and $Y_0$ which is articulated in the next subsection, the linear combination can be used to construct the point estimator and CI for the 0-th QTE.
2.5.3 A robust confidence interval

The inference methods for intermediate and extreme QTE estimators are different. This raises the practical issue of how to choose the inference method in a given dataset with a small but given quantile index. Note that for \( a \in (0, 1) \), any two-sided \((1 - a)\)-th CI can be written as

\[
\text{CI} = \left( \hat{q}(\tau_n) - \tilde{C}_{1-a/2}(\tau_n), \hat{q}(\tau_n) - \tilde{C}_{a/2}(\tau_n) \right)
\] (2.5.4)

where \( \tilde{C}_a(\tau_n) \) is the critical value. However, the choice of \( \tilde{C}_a(\tau_n) \) depends on the order of \( \tau_n \).

Ideally, for extreme-order quantile index,

\[
\tilde{C}_a(\tau_n) = \tilde{C}^{bn}_a(\tau_n) := \tilde{C}_a(\tau_n)/\hat{\alpha}_n
\]

where \( \tilde{C}_a(\tau_n) \) is the critical value computed by a \( b \) out of \( n \) bootstrap procedure for \( \tau_n \). For the intermediate and regular order quantile indices, \( \tilde{C}_a(\tau_n) = \tilde{C}^{mn}_a(\tau_n) \) where \( \tilde{C}^{mn}_a(\tau_n) \) is the critical value computed by a standard bootstrap procedure. But in practice, it is impossible to determine the order of any quantile index because researchers only have one dataset. Hence the ideal procedure is not feasible.

The same problem is encountered in Andrews and Cheng (2012) in which the model can be either weakly, semi-strongly, or strongly identified. What they propose is an identification-category-selection (ICS) procedure. Similarly, here, I propose an order-category-selection (OCS) procedure based on the quantile index of interest and construct a robust CI.

Let \( \tau_{n,1} := \min\left( \frac{40}{n}, \frac{0.2b}{mn} \right) \), \( \tau_{n,2} = \frac{b}{n \sqrt{\log(n)}} \), and for any \( a \in (0, 1) \),

\[
\tilde{C}^{df}_{a/2}(\tau_n) = \max(\tilde{C}^{bn}_{a/2}(\tau_n), \tilde{C}^{mn}_{a/2}(\tau_n)) \quad \text{and} \quad \tilde{C}^{df}_{1-a/2}(\tau_n) = \min(\tilde{C}^{bn}_{1-a/2}(\tau_n), \tilde{C}^{mn}_{1-a/2}(\tau_n)).
\]
The robust CI is constructed based on a hybrid critical value \( \tilde{C}_a^h(\tau_n) \) defined as follows.

\[
\begin{align*}
\tilde{C}_a^h(\tau_n) &= \begin{cases} 
\tilde{C}_a^{bn}(\tau_n) & \text{if } \tau_n \leq \tau_{n,1} \\
\tilde{C}_a^{df}(\tau_n) & \text{if } \tau_n \in (\tau_{n,1}, \tau_{n,2}) \\
\tilde{C}_a^{mn}(\tau_n) & \text{if } \tau_n \geq \tau_{n,2}.
\end{cases}
\end{align*}
\]

For \( n \) large enough, \( \tau_{n,1} = \frac{40}{n} \). If \( \tau \leq \tau_{n,1} \), \( n \tau \leq 40 < \infty \). For such \( \tau \), it is expected that the extreme-order asymptotic distribution can approximate the finite distribution of the \( \tau \)-th QTE estimator better than the standard normal distribution. In this case, the hybrid CI equals BN-CI. The choice of two constants in \( \tau_{n,1} \) are discussed in Section 4.5.

For \( \tau_{n,2} \), if \( \tau \geq \tau_{n,2} \),

\[
\tau n \geq \frac{b}{\sqrt{\log(n)}} \rightarrow \infty
\]

because \( b \rightarrow \infty \) polynomially in \( n \). For such \( \tau \), it is expected that the finite sample distribution of the \( \tau \)-th QTE estimator is well approximated by the intermediate or regular order quantile asymptotic distribution. In both cases, the standard bootstrap CI is consistent. In addition, \( \tau \geq \tau_{n,2} \) also implies that

\[
\tau_b := \frac{n \tau}{b} \geq \frac{1}{\sqrt{\log(n)}}.
\]

It means that the quantile index \( \tau_b \) used in computing the \( b \) out of \( n \) CI is not small. Thus to view \( \tau_b \) in practice as an extremal quantile is not appropriate and the asymptotic properties of the \( b \) out of \( n \) CI computed using \( \tau_b \) are not effective. In this case, the hybrid CI equals the standard bootstrap CI.

When \( \tau \in (\tau_{n,1}, \tau_{n,2}) \), whether normal or EV approximation works better is not clear. In this case, the hybrid CI uses the least favorable critical value which is conservative.
The OCR procedure is different from the ICS procedure used in Andrews and Cheng (2012) because here I have two thresholds and when the quantile index is less than the first threshold, the asymptotic size is exact, while in Andrews and Cheng (2012), they only have one threshold and when the strength of identification is less than the threshold, their asymptotic size is conservative.

Let

\[ \Gamma_{ex} := \left\{ \{ \tau_n \}_{n \geq 1} : \tau_n \to 0, n\tau_n \to k \in (0, \infty), k \right\}
\]

\[ \Gamma_{int} := \left\{ \{ \tau_n \}_{n \geq 1} : \tau_n \to 0, n\tau_n \to \infty \right\}, \]

and

\[ \Gamma_{reg} := \left\{ \{ \tau_n \}_{n \geq 1} : \tau_n = k \in (0, 1) \right\} \]

denote the collections of extreme, intermediate, and regular order sequences of quantile indices. Next Theorem shows that the new CI is robust over \( \Gamma := \Gamma_{ex} \cup \Gamma_{int} \cup \Gamma_{reg} \).

**Theorem 2.5.3.** Assumptions 1, 3–5, and 7–8 hold. Subsample size \( b \to \infty \) polynomially in \( n \) and \( \frac{b}{n} \to 0 \). The standard bootstrap inference is consistent for regular quantile indices. Then, for any \( a \in (0, 1) \),

\[ \inf_{\{ \tau_n \}_{n \geq 1} \in \Gamma} \lim_{n \to \infty} \mathbb{P} \left( \hat{q}(\tau_n) \in \left( \hat{\tau}(\tau_n) - \hat{C}_{\frac{1}{2}}^{h} (\tau_n), \hat{\tau}(\tau_n) - \hat{C}_{\frac{3}{2}}^{h} (\tau_n) \right) \right) = 1 - a. \]

Unlike Andrews and Cheng (2012) in which the parameters and thus the d.g.p.’s are drifting, in my case, the d.g.p. is fixed and the quantile index is drifting. So the above result mainly focus on the robustness of CI’s over different categories of quantile orders but does not speak to the uniformity over different d.g.p.’s.
2.5.4 Inference theory for the 0-th QTE

This section constructs a consistent CI for the 0-th QTE when the lower boundaries of \( Y_1 \) and \( Y_0 \) are bounded. The estimator for the 0-th QTE is a linear combination of extreme-order QTE estimators with a set of carefully chosen weights. For inference, the same procedure of the inference method proposed for the extreme QTE in Section 2.5.3 can be directly applied.

I use a linear combination of extreme QTE estimators to infer the 0-th QTE so that the estimation bias cancels out. To see the source of bias, I first recall that, when the lower end point is bounded and Assumption 15 holds, the tail is Type 3. This implies that \( q^*_j = q_j(0) \) and \( \beta_{n,j} = 0 \). Hence I have

\[
\hat{q}(\tau_n) - (q_1(0) - q_0(0)) = \hat{q}(\tau_n) - q(\tau_n) + \frac{k^{-\xi_1} + o(1)}{\alpha_{1,n}} - \frac{k^{-\xi_0} + o(1)}{\alpha_{0,n}}. \tag{2.5.5}
\]

I can approximate the critical value of the asymptotic distribution for \( \hat{q}(\tau_n) - q(\tau_n) \) based on the procedure after Proposition 2.5.1. The second term on the RHS of (B.4.7) is the bias caused by the fact that the parameter of interest is \( q(0) \), instead of \( q(\tau_n) \).

To get rid of this bias, I propose a feasible estimator \( \hat{q}(0) := \sum_{l=1}^L \hat{r}_l \hat{q}(\tau_{n,l}) \) in which the weights \( \{\hat{r}_l\}_{l=1}^L \) solve the following system of equations:

\[
\sum_{l=1}^L \hat{r}_l = 1, \quad \sum_{l=1}^L \hat{r}_l k_l^{-\xi_1} = 0, \quad \sum_{l=1}^L \hat{r}_l k_l^{-\xi_0} = 0. \tag{2.5.6}
\]

Here, \((\hat{\xi}_0, \hat{\xi}_1)\), the consistent estimators of \((\xi_0, \xi_1)\), can be computed by Theorem 2.3.4.

To implement, I compute \( \hat{q}(0) \) using only three different values of \( \tau_{n,l} \), that is, \( L = 3 \). The reason is twofold: (1) I do not have a selection rule for choosing among solutions of weights that satisfies (2.5.6) if the solution is not unique, and (2) by
fixing the upper and lower bound \( \tau_{n,1} \) and \( \tau_{n,L} \), the more quantile indices I use, the higher the weights, which will widen the implied CI.

**Proposition 2.5.2.** Let \( \hat{\xi}_j \) be consistent estimates of \( \xi_j \) for \( j = 0, 1, L = 3 \), \( \hat{r}_1, \hat{r}_2, \hat{r}_3 \) be computed as in (2.5.6), \( \hat{q}(0) := \sum_{l=1}^{L} \hat{r}_l \hat{q}(\tau_{n,l}) \), and \( \hat{C}_a \) be computed as in the procedure after Proposition 2.5.1. If the assumptions in Theorem 2.4.2 hold and \( q_j(0) \) is bounded for \( j = 0, 1 \), then

\[
\lim_{n \to \infty} P \left( \frac{\hat{q}(0) - \hat{C}_1 - a/2}{\hat{\alpha}_n} \leq q(0) \leq \frac{\hat{q}(0) - \hat{C}_a/2}{\hat{\alpha}_n} \right) = 1 - a.
\]

There are two alternative methods by which to infer the 0-th QTE, each of which has its own restrictions. The first alternative is to analytically compute \( \frac{\alpha_{-0}}{\alpha_{-1,n}} - \frac{\alpha_{-0}}{\alpha_{0,n}} \), the leading term of the bias in (B.4.7). This requires the estimation of the infeasible convergence rate \( \alpha_{j,n} \). However, computing an estimator \( \hat{\alpha}_{j,n} \) of \( \alpha_{j,n} \) such that \( \frac{\hat{\alpha}_{j,n}}{\alpha_{j,n}} \to 1 \) is harder than simply estimating the EV index \( \xi_j \). Usually, in order to compute \( \hat{\alpha}_{j,n} \), distributional assumptions, such as \( \alpha_{j,n} = C_j n^{\delta_j} \) for some constant \( C_j \), are imposed. See, for example, the discussion in Chernozhukov and Fernández-Val (2011) on the distributional assumption and Bertail et al. (1999) on the point of unknown convergence rate in subsampling inferences. These assumptions are not needed in Proposition 2.5.2.

The second alternative is to rely on asymptotics to ensure that the bias is asymptotically negligible and small in the finite sample. To be more specific, combining Theorems 2.4.1 and 2.4.2, it is clear that for \( \tau_n n \to k \),

\[
\hat{\alpha}_n(\hat{q}(\tau_n) - q(0))
\]

converges weakly to a non-degenerate limiting distribution. I can then approximate the critical value of the limiting distribution by computing

\[
\hat{Z}_n^*(k) := \hat{\alpha}_n^*(\hat{q}_n^*(\tau_n) - \hat{q}(\tau_n))
\]
for $\tau_b = \tau_n n$. Comparing $\hat{Z}_n^*(k)$ with $\hat{Z}_n^{cs}(k)$ in (2.5.2), the only difference is that the subsample estimator $\hat{q}^*_b(\tau_b)$ is now centered by $\hat{q}(\tau_n) := \hat{q}_1(\tau_n) - \hat{q}_0(\tau_n)$, the full sample QTE estimator at $\tau_n$, instead of $\hat{q}(\tau_b)$. The reason is that for the subsample, $\hat{q}(\tau_b)$ and $\hat{q}(\tau_n)$ can be viewed as proxies for $q(\tau_b)$ and $q(0)$, respectively. This corresponds to the two different parameters of interest (i.e., $q(\tau_b)$ and $q(0)$) in these two situations. Then after I obtain an estimator of the critical value of the limiting distribution of $\hat{Z}_n^*(k)$ by a similar $b$ out of $n$ bootstrap procedure, I can construct a median-unbiased estimator and a consistent CI for $q(0)$. This method is passive because econometricians have no control on the magnitude of the bias in a finite sample. This implies that the properties of the implied CI in finite samples are sensitive to both the choice of $k = \tau_n n$ and the subsample size $b$. Therefore, the passive method is less robust than the one proposed in Proposition 2.5.2.

2.6 Simulations

This section verifies the usefulness of the asymptotic theory and the performance of BN-CI through a simulation study. It shows that the BN-CI is consistent, insensitive to the choice of subsample size $b$, while the standard bootstrap CI (NN-CI) undercovers substantially.

2.6.1 Limiting distributions

I first verify the asymptotic distributions of $\hat{q}_1(\tau_n)$ established in Section 2.4. Figure 2.5 plots the normalized sample distribution of $\hat{q}_1(\tau_n)$ against its limiting distribution established in Theorem 2.4.1 with four different boundary structures: single minimizer, finite minimizers, continuum minimizers, and mixture minimizers. Since the plots are all close to the diagonal line, it implies that the new asymptotic distributions based established in Theorem 2.4.1 approximate the finite sample distributions very well.
To compute the sample estimator, I generate random samples with size 1,000 and repeat both the estimation and the minimization of the asymptotic objective function 400 times. $k := \tau_n n$ is set to 5. The propensity score is estimated in a sieve approach by fitting a series logistic model with ordinary polynomial basis to the fourth order.

Figure 2.6 plots the exact same estimators against the standard normal distribution. The plots are all non-linear, which indicates that the shape of the finite sample distributions is not normal. Any inference method based asymptotic normality will fail to produce a consistent CI.

2.6.2 Inference for the extreme QTE

In this section, I investigate the finite sample performance of the $b$ out of $n$ bootstrap CI with replacement (BN-CI) for the extreme QTE, as proposed in Proposition 2.5.1.
In the simulation, \( n = 5,000 \), \( k \) is fixed at \((5, 10, 20, 40)\), and the corresponding quantile indices are \( \tau_n = (0.1\%, 0.2\%, 0.4\%, 0.8\%) \). The subsample size used in Table 2.1 and Figure 2.7 is 1,000. In Table 2.1, 2.2, Figure 2.7, and Figure 2.8, I consider four simulation designs corresponding to four different boundary conditions as in the previous subsection for both \( Y_1 \) and \( Y_0 \): (1) single minimizer, (2) multiple minimizers, (3) continuum minimizers, and (4) mixture minimizers. Table 2.1 and 2.2 report the coverages of BN-CI and NN-CI, respectively. The number in the parentheses is the median length of the CI. Figure 2.7 plots the coverages of BN-CI and NN-CI against \( \tau \) for \( \tau \in [0.1\%, 2\%] \). Figure 2.8 plots the coverage of BN-CI against \( b \) for \( b \in [500, 1,500] \).

**Table 2.1: Coverage of 95% \( b \) out of \( n \) bootstrap CI, sample size = 5,000**

<table>
<thead>
<tr>
<th>( \tau_n ) = 0.1%, ( k = 5 )</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>( \tau_n ) = 0.2%, ( k = 10 )</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.941</td>
<td>0.936</td>
<td>0.941</td>
<td>0.940</td>
<td>(1)</td>
<td>0.955</td>
<td>0.948</td>
<td>0.95</td>
<td>0.949</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.023)</td>
<td>(0.017)</td>
<td>(0.022)</td>
<td></td>
<td>(0.029)</td>
<td>(0.025)</td>
<td>(0.019)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>(2)</td>
<td>0.948</td>
<td>0.944</td>
<td>0.943</td>
<td>0.941</td>
<td>(2)</td>
<td>0.953</td>
<td>0.942</td>
<td>0.961</td>
<td>0.949</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.019)</td>
<td>(0.012)</td>
<td>(0.020)</td>
<td></td>
<td>(0.028)</td>
<td>(0.021)</td>
<td>(0.014)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>(3)</td>
<td>0.957</td>
<td>0.948</td>
<td>0.947</td>
<td>0.939</td>
<td>(3)</td>
<td>0.959</td>
<td>0.957</td>
<td>0.966</td>
<td>0.956</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.018)</td>
<td>(0.006)</td>
<td>(0.012)</td>
<td></td>
<td>(0.026)</td>
<td>(0.020)</td>
<td>(0.007)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>(4)</td>
<td>0.954</td>
<td>0.938</td>
<td>0.940</td>
<td>0.935</td>
<td>(4)</td>
<td>0.959</td>
<td>0.949</td>
<td>0.941</td>
<td>0.950</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.018)</td>
<td>(0.014)</td>
<td>(0.015)</td>
<td></td>
<td>(0.028)</td>
<td>(0.021)</td>
<td>(0.017)</td>
<td>(0.019)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \tau_n ) = 0.4%, ( k = 20 )</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>( \tau_n ) = 0.8%, ( k = 40 )</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.968</td>
<td>0.962</td>
<td>0.949</td>
<td>0.956</td>
<td>(1)</td>
<td>0.979</td>
<td>0.974</td>
<td>0.977</td>
<td>0.967</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.026)</td>
<td>(0.020)</td>
<td>(0.027)</td>
<td></td>
<td>(0.037)</td>
<td>(0.032)</td>
<td>(0.026)</td>
<td>(0.033)</td>
</tr>
<tr>
<td>(2)</td>
<td>0.956</td>
<td>0.967</td>
<td>0.968</td>
<td>0.953</td>
<td>(2)</td>
<td>0.969</td>
<td>0.968</td>
<td>0.965</td>
<td>0.953</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.023)</td>
<td>(0.015)</td>
<td>(0.027)</td>
<td></td>
<td>(0.034)</td>
<td>(0.027)</td>
<td>(0.020)</td>
<td>(0.033)</td>
</tr>
<tr>
<td>(3)</td>
<td>0.960</td>
<td>0.960</td>
<td>0.951</td>
<td>0.947</td>
<td>(3)</td>
<td>0.963</td>
<td>0.966</td>
<td>0.968</td>
<td>0.970</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.022)</td>
<td>(0.008)</td>
<td>(0.016)</td>
<td></td>
<td>(0.031)</td>
<td>(0.025)</td>
<td>(0.011)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>(4)</td>
<td>0.962</td>
<td>0.949</td>
<td>0.939</td>
<td>0.945</td>
<td>(4)</td>
<td>0.983</td>
<td>0.972</td>
<td>0.972</td>
<td>0.973</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.022)</td>
<td>(0.018)</td>
<td>(0.021)</td>
<td></td>
<td>(0.041)</td>
<td>(0.033)</td>
<td>(0.027)</td>
<td>(0.031)</td>
</tr>
</tbody>
</table>
Table 2.2: Coverage of 95% $n$ out of $n$ bootstrap CI, sample size = 5,000

<table>
<thead>
<tr>
<th>$\tau_n$</th>
<th>0.1%, $k = 5$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>$\tau_n$</th>
<th>0.2%, $k = 10$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.835</td>
<td>0.842</td>
<td>0.829</td>
<td>0.825</td>
<td>(1)</td>
<td>0.869</td>
<td>0.855</td>
<td>0.853</td>
<td>0.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.019)</td>
<td>(0.014)</td>
<td>(0.017)</td>
<td></td>
<td>(0.024)</td>
<td>(0.020)</td>
<td>(0.016)</td>
<td>(0.019)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>0.830</td>
<td>0.835</td>
<td>0.850</td>
<td>0.790</td>
<td>(2)</td>
<td>0.861</td>
<td>0.848</td>
<td>0.875</td>
<td>0.841</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.016)</td>
<td>(0.010)</td>
<td>(0.015)</td>
<td></td>
<td>(0.021)</td>
<td>(0.017)</td>
<td>(0.012)</td>
<td>(0.017)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>0.768</td>
<td>0.783</td>
<td>0.844</td>
<td>0.775</td>
<td>(3)</td>
<td>0.828</td>
<td>0.824</td>
<td>0.873</td>
<td>0.830</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.013)</td>
<td>(0.004)</td>
<td>(0.009)</td>
<td></td>
<td>(0.019)</td>
<td>(0.015)</td>
<td>(0.005)</td>
<td>(0.011)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>0.793</td>
<td>0.835</td>
<td>0.852</td>
<td>0.819</td>
<td>(4)</td>
<td>0.846</td>
<td>0.865</td>
<td>0.858</td>
<td>0.863</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.014)</td>
<td>(0.011)</td>
<td>(0.012)</td>
<td></td>
<td>(0.020)</td>
<td>(0.016)</td>
<td>(0.013)</td>
<td>(0.014)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\tau_n$</th>
<th>0.4%, $k = 20$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>$\tau_n$</th>
<th>0.8%, $k = 40$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.891</td>
<td>0.891</td>
<td>0.882</td>
<td>0.891</td>
<td>(1)</td>
<td>0.903</td>
<td>0.919</td>
<td>0.890</td>
<td>0.892</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.022)</td>
<td>(0.018)</td>
<td>(0.021)</td>
<td></td>
<td>(0.027)</td>
<td>(0.024)</td>
<td>(0.020)</td>
<td>(0.025)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>0.878</td>
<td>0.898</td>
<td>0.906</td>
<td>0.864</td>
<td>(2)</td>
<td>0.889</td>
<td>0.909</td>
<td>0.903</td>
<td>0.877</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.019)</td>
<td>(0.014)</td>
<td>(0.020)</td>
<td></td>
<td>(0.023)</td>
<td>(0.020)</td>
<td>(0.015)</td>
<td>(0.023)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>0.871</td>
<td>0.860</td>
<td>0.882</td>
<td>0.865</td>
<td>(3)</td>
<td>0.879</td>
<td>0.881</td>
<td>0.903</td>
<td>0.885</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.016)</td>
<td>(0.006)</td>
<td>(0.012)</td>
<td></td>
<td>(0.020)</td>
<td>(0.017)</td>
<td>(0.007)</td>
<td>(0.015)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>0.880</td>
<td>0.879</td>
<td>0.880</td>
<td>0.904</td>
<td>(4)</td>
<td>0.899</td>
<td>0.881</td>
<td>0.894</td>
<td>0.912</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.018)</td>
<td>(0.015)</td>
<td>(0.017)</td>
<td></td>
<td>(0.023)</td>
<td>(0.021)</td>
<td>(0.019)</td>
<td>(0.021)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.1 and 2.2 illustrate that the standard bootstrap CI undercovers as much as 18.2 absolute percentage points while the BN-CI’s coverage is very close to the nominal 95% when $\tau$ is less than 2% or correspondingly, $k := \tau n \leq 40$. In addition, the length of the BN-CI is larger but still comparable to one with the standard bootstrap CI, which ensures the practical value of BN-CI.

Next, I explain the choice of two constants in $\tau_{n,1}$. In general

$$\tau_{n,1} = \min \left( \frac{C_1}{n}, \frac{C_2 b}{mn} \right),$$

where $C_1$ and $C_2$ are two positive constant. When $\tau \leq \tau_{n,1}$, I suggest researchers to use the BN-CI to conduct inferences.

For $C_1$, if $k := \tau n$ is large, the approximation error from estimating the propensity score will contaminate the asymptotic approximation. This inspires the requirement
that \( n\tau \leq C_1 \). Chernozhukov (2005a) and Chernozhukov and Fernández-Val (2011) suggest to use \( C_1 \in [40, 80] \). To be cautious, I choose \( C_1 = 40 \).

Second, the EV-law asymptotic approximation is only valid in the subsample with subsample size \( b \) if the quantile index used in the subsample, \( m\tau_b := \frac{mk}{b} = \frac{m\tau n}{b} \), is close to zero. This inspires the second requirement that

\[
m\tau_b \leq C_2.
\]

Based on the simulations, the quantile index \( m\tau_b \) is small enough if it is less than \( C_2 = 0.2 \). Combining these two requirements, I obtain \( \tau_{n,1} \).

The next figure shows that when the quantile index is less than the threshold, the BN-CI has an accurate coverage while the standard bootstrap CI (NN-CI) undercovers substantially. As the quantile index increases, BN-CI usually overcovers, which means that the BN-CI is conservative, while the NN-CI still undercovers, but the coverage gradually converges to the nominal rate.
Each \((i, j)\)-th subplot represents the \((i, j)\)-th model. The dashed line is the coverage of BN-CI with \(b = 1,000\) and \(n = 5,000\) for quantile index \(\tau \in [0.1\%, 2\%]\). The dotted line is the coverage of NN-CI. The horizontal dotted dashed line is the 95\% nominal coverage rate, and the vertical dotted dashed line is \(\tau = \min(\frac{40}{n}, \frac{0.2b}{mn})\).

\textbf{Figure 2.7: Coverage across quantiles}

The next figure shows that the BN-CI is insensitive to the choice of subsample size \(b\) over a reasonable range. The choice of subsample size \(b\) is an important tuning parameter. Since I can estimate the EV-indices for \(Y_0\) and \(Y_1\) by Theorem 2.3.4, it is possible to recover the leading term of the convergence rates for \(\hat{q}_1(\tau_n)\) and \(\hat{q}_0(\tau_n)\) when \(\tau_n\) is extreme. This implies that it is possible to compute the "almost optimal"\(^6\) rate of the subsample size \(b\) if I have a similar Edgeworth expansion of the extreme QTE estimator around the new asymptotic distribution established in Theorem 2.5.2. This is left as a useful research direction.

\(^6\) For example, the convergence rate for \(\hat{q}_j(\tau_n)\) can be \(\log(\tau_n)\tau_n^{-\xi_j}\) where the leading term is \(\tau_n^{-\xi_j}\). By estimating the EV-index \(\xi_j\), I can recover the leading term but cannot recover the \(\log(\tau_n)\) term. If the rate of \(b\) is calculated only using the leading term \(\tau_n^{-\xi_j}\), it is optimal up to a \(\log(n)\) term.
Here I only report the results for sample size 5,000. The same simulation designs with sample size 300 and 1,000 can be found in the Appendix A.1.3. All the above findings still hold. In Appendix A.1.3, I also show the mean bias (bias), root mean square error (rMSE), median bias (mbias), and mean absolute error (MAE) of the median-unbiased point estimator for small, moderate and large sample. The performance of the median-unbiased point estimator is satisfying in all samples.

2.6.3 The robust confidence interval

The next figure shows the finite sample performance of the robust CI proposed in Section 2.5.3. To implement, the full sample size and subsample size are $n = 5,000$ and $b = 1,000$, respectively. $Y_1$ has a single minimizer and $Y_0$ has continuum minimizers. The quantile index $\tau \in [0.1\%, 8\%]$. For computing $\tilde{C}^{bn}_a(\tau)$, when $\tau \leq 2\%$
or equivalently, $k := \tau n \leq 100$, I set the spacing parameter $m = 2$ and $k_l' = 10$. When $\tau > 2\%$, I set $m = 1.2$ and $k_l' = 20$.

When $\tau \leq \tau_{n,1}$ or $\tau \geq \tau_{n,2}$, the coverage is close to the 95% nominal rate while when $\tau \in (\tau_{n,1}, \tau_{n,2})$, the CI overcovers and thus is conservative. All sixteen models exhibit this same pattern. For details, please see Section A.1.3.

2.6.4 Inference for the 0-th QTE

In this section, I report the finite sample performance of the inference method for the 0-th QTE established in Section 2.5.4. Table 2.3 shows the coverage and median length of BN-CI. Figure 2.10 plots the coverage of BN-CI against the subsample size $b$ for $b \in [500, 1,000]$. I find that the BN-CI is consistent and insensitive to the choice of subsample size $b$ over a reasonable range. Here again I only focus on $n = 5,000$. The same simulation with $n = 300,1,000$ can be found in the appendix. All the

\[ k_l' \text{ is used to compute the normalizing factor } \hat{\alpha}_n. \]
above findings still hold.

There are two additional issues in inferring the 0-th QTE. The first issue is that I use three extreme QTE estimators with \( k = (5, 17.5, 30) \) to compute the linear combination. The choice of \( k \) invokes two concerns. First, as illustrated above, the rule of thumb for \( k = \tau n \) is \( k \leq \min(40, \frac{0.26}{m}) \). Second, the space among \( k \)'s must not be narrow, otherwise the weights will be large in absolute value, which will widen the CI. The second issue is that I estimate the EV indices following Theorem 2.3.4 with \( R = 2, m = 2, l = 2 \), and equal weights. The set of quantile indices I use to compute the EV indices are \( \tau_n = (0.002, 0.004, \ldots, 0.01) \). Then for \( j = 0, 1 \), the two EV index estimators used to compute the weights \( \hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3 \) are the median of the estimators computed using each of the quantile indices for \( j = 0 \) and 1, respectively.

The rest of the simulation details are the same as the ones in the previous subsection. The subsample size for Table 2.3 is 1,000.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.960</td>
<td>0.955</td>
<td>0.951</td>
<td>0.948</td>
</tr>
<tr>
<td></td>
<td>(0.118)</td>
<td>(0.086)</td>
<td>(0.058)</td>
<td>(0.077)</td>
</tr>
<tr>
<td>(2)</td>
<td>0.954</td>
<td>0.947</td>
<td>0.961</td>
<td>0.951</td>
</tr>
<tr>
<td></td>
<td>(0.105)</td>
<td>(0.073)</td>
<td>(0.042)</td>
<td>(0.072)</td>
</tr>
<tr>
<td>(3)</td>
<td>0.957</td>
<td>0.952</td>
<td>0.955</td>
<td>0.948</td>
</tr>
<tr>
<td></td>
<td>(0.082)</td>
<td>(0.056)</td>
<td>(0.017)</td>
<td>(0.036)</td>
</tr>
<tr>
<td>(4)</td>
<td>0.956</td>
<td>0.919</td>
<td>0.951</td>
<td>0.953</td>
</tr>
<tr>
<td></td>
<td>(0.086)</td>
<td>(0.053)</td>
<td>(0.039)</td>
<td>(0.044)</td>
</tr>
</tbody>
</table>
The solid line is the coverage for $b$ out of $n$ bootstrap CI at \( \tau = 0 \) in which $b \in [500, 1,000]$.

**Figure 2.10:** Coverage across subsample size

### 2.7 Empirical application

#### 2.7.1 Effect of maternal status on extremely low birth weights

This section applies the inference method I develop in this paper to examine the effect of maternal smoking on the lower tail of babies’ birth weights. I first investigate the QTE of maternal smoking on the lower end and the lower boundary of birth weight distribution based on BN-CI. Although I cannot reject the null hypothesis that the maternal smoking has no effect on the lower tail of the birth weights at the 90% confidence level, I find that the BN-CI is two times wider than the NN-CI. This indicates that the NN-CI of the effect of maternal smoking potentially undercovers or over-rejects.

The lower tail of the birth weight distribution reflects the severity of the adverse birth outcome, which is the main research interest in health economics. Adverse birth outcomes, particularly the low birth weight, are one of the leading causes of
infant mortality, a main concern of public health research. In addition, adverse birth outcomes result in large economic costs in terms of both direct newborn care costs and long-term developmental costs such as delayed entry into kindergarten, repeated grades and the consequent labor market outcomes. For literature on maternal smoking and birth weights, see, for example, Abrevaya (2002), Abrevaya (2006), Abrevaya and Dahl (2008), Chernozhukov and Fernández-Val (2011), Evans and Lien (2005), Evans and Ringel (1999) Permutt and Hebel (1989), Rosenzweig and Wolpin (1991), and the references therein.

Despite the large literature on the effect of maternal smoking on birth weights, there is no consensus on its magnitude. It is found that the negative effect of maternal smoking is a decreased weight of about 189-600 grams\(^8\) in various research papers by using different estimation tools and data. See Abrevaya (2006) for a summary. But in order to draw these conclusions, empirical researchers usually consider small but regular quantile estimates or subsamples of low-weight infants and refer to the asymptotic normality to draw inferences. The only exception is Chernozhukov and Fernández-Val (2011), who looked at extremely low birth weight and referred to the EV distribution to draw inferences. Figure 8 of Chernozhukov and Fernández-Val (2011) shows that the extremal quantile regression coefficient of maternal smoking is close to zero and statistically insignificant.

I estimate the QTE of maternal smoking on extremely low birth weight infants. The QTE is distinct from the linear regression coefficient of smoking status estimated in Chernozhukov and Fernández-Val (2011) in three aspects. First, the extreme QTE is an unconditional parameter while the regression coefficient is a conditional one. The unconditional extreme QTE estimated here empirically differs from the linear regression coefficient because the conditional quantile is heterogeneous as shown in

\(^{8}\) The average birth weight for an infant is about 3,400 grams.
Figure 8 and 9 in Chernozhukov and Fernández-Val (2011). To recover the unconditional QTE from a conditional coefficient is also nontrivial because inverse c.d.f. is a nonlinear operator. Furthermore, the unconditional extreme QTE is an aggregate measure which is policy-relevant. It avoids modeling the conditional quantile, which makes the estimator robust to misspecifications. For the second difference, I estimate the parameter under the causal inference framework by controlling confounders while for the linear regression coefficient to have a causal interpretation, one must assume exogeneity of smoking status. Last, I also estimate the exact 0-th unconditional QTE which is new to the literature.

I use the same dataset as in Chernozhukov and Fernández-Val (2011). It was collected based on June 1997 Detailed Natality Data published by the National Center for Health Statistics and has been previously investigated by Abrevaya (2002) and Koenker and Hallock (2001). I concentrate on African American mothers only, with 31,912 observations, because Figure 7 of Chernozhukov and Fernández-Val (2011) shows that low birth weights for black mothers have a heavy lower tail. Economically, it indicates a severe adverse birth outcome which is the main target of this analysis. Theoretically, the heavy lower tail of the birth weights distribution is consistent with Assumption 15(3), which is the key to conducting the $b$ out of $n$ bootstrap inference for the extreme QTE.

To fit the notation in the paper, let $D$ be an indicator of maternal smoking. The observed outcome variable $Y$ is birth weight measured in grams, while $Y = DY_1 + (1 - D)Y_0$ where $Y_1$ is the infant’s potential birth weight when the mother smokes and $Y_0$ is the infant’s potential birth weight when the mother does not smoke. Covariates $X$ are demographic variables which include mother’s age, mother’s education level\textsuperscript{9}, an indicator of whether the mother had parental care visit in the first

\textsuperscript{9} The education level equals 0 if the mother has less than a high school education, 1 if she completed high school, 2 if she obtained some college education, and 3 if she graduated from college.
and second trimester, mother’s marriage status, the infant’s sex, and mother’s weight gain during pregnancy. The key unconfoundedness assumption in this context means that, maternal smoking is independent from the potential birth weights conditional on all the demographic variables.

Following the experience collected from Section 4.5, I set the subsample size to 3,000 and repeat the \( b \) out of \( n \) bootstrap with replacement 20,000 times. Also, I nonparametrically estimate the propensity by fitting a series logistic model with a set of second-order polynomial basis, and the spacing parameter \( m \) is set to 2.\(^{10}\) When computing the 0-th QTE, I use a linear combination of extreme-order estimates with \( k = (5, 20, 40) \). A set of estimators of EV index are computed following Theorem 2.3.4 with \( R = 2, l = 2 \) and \( \tau_n = (0.0005, 0.001, 0.0015, 0.002, 0.0025) \). The final EV index estimators used are the median of the five estimators for \( Y_0 \) and \( Y_1 \), respectively.

Table 2.4 reports the median-unbiased point estimates and the CI for the extreme QTE of maternal smoking. In all quantile indices, I cannot reject that maternal smoking has no negative impact on either extremal quantile or the lower bound of infants’ birth weights under 90% confidence level. On the other hand, the BN-CI is more than two times wider than the standard bootstrap CI. This indicates that the standard bootstrap CI potentially undercovers which is consistent with the simulation study. Last, the median-unbiased estimator for the 0-th QTE implies that if a pregnant mother smokes, with 50% probability, her child’s lowest possible birth weight is 137.32 grams lighter than it would be if she did not smoke.

Although estimating the extreme QTE is one step forward in the direction of causal inference, the existence of other confounders that can jeopardize the selection on observables is possible. For example, mothers who smoke during pregnancy are

\(^{10}\) Here I implicitly assume that the sufficient condition for the spacing parameter in Lemma A.1.7 holds. In practice, neither the full sample nor any subsample estimation encounters the zero denominator error. Hence \( m = 2 \) behaves well in this data analysis.
more likely to adopt other behaviors (drinking, poor nutritional intake, etc.) that could have a negative impact on birth weight. Evans and Lien (2005) and Evans and Ringel (1999) address this problem by using large cigarette taxation change as an instrumental variable (IV) for maternal smoking. Extending the current theory to incorporate IV and conduct inference for the extremal QTE for the compliers would be a useful research direction.

2.7.2 Effect of minority status on college preparation index

This section considers the effect of minority status on the college preparation index (CPI) for low-scoring students with equivalent family backgrounds. This gap can be viewed as a measure of affirmative action in colleges’ admission selections. As described in Arcidiacono et al. (2014), in 1996, the voters of California approved Prop 209 which stipulates that: “The state shall not discriminate against, or grant preferential treatment to, any individual or group on the basis of race, sex, color, ethnicity, or national origin in the operation of public employment, public education, or public contracting.” The proposition took effect in 1998. By computing and comparing the extreme QTE of minority status on CPI for students admitted by at least one campus of University of California both prior and post Prop 209, I investigate that how the implementation of Prop 209 influences colleges’ admission
Students with low academic credentials are more likely to be affected by the change of admission selection criteria. If a college’s admission is purely meritocratic, Proposition 1 of Bhattacharya et al. (2014) shows that the optimal admission protocol is a simple threshold-crossing form. Given the population of enrolled students, the threshold can be identified as the lower boundary of the CPI distribution which is the zero-th quantile. The gap of zero-th quantile of the distributions of CPI for minority and majority students can be viewed as a measure of the magnitude of racial preference in college admission. In contrast, to use the average gap of CPI (ATE) to measure racial preference may be problematic because two distributions can have the same lower boundaries but different first moments. In reality, the admission criteria in U.S. is multidimensional and not purely meritocratic. Therefore, there is no simple threshold for CPI. However, following the guidance of the simple model, it is reasonable to say that low-scoring students are more likely to be affected by the policy change on racial preferences in colleges’ admission selections. This implies that the extremal QTE of minority status on CPI measures the magnitude of racial preference in college admission.

Since my analysis focuses on the tail population of the CPI distribution, it is similar to Bhattacharya et al. (2014) which focuses on the marginal admits but is in contrast with many other studies which focus on average pre-admission test-scores (e.g. Zimdars et al. (2009)) or average post-admission test-scores (e.g. Keith et al. (1985), Kane (1998), and Sackett et al. (2009)). For historical perspective on selectivity in US college admission, see Hoxby (2009).

I use the same data as in Arcidiacono et al. (2015), the University of California Office of the President (UCOP) data for minority and non-minority students who first enrolled at one of the UC campuses between 1995 through 1997 during which
racial preference in admission was legal. The data for each UC campuses consist of all their admitted students. The outcome variable $Y$ is normalized CPI.\textsuperscript{11} The treatment status $D$ is the indicator of under-represented minority groups in the dataset. $X$ are two family background variables: family income percentage and two parents’ highest education degree. Minority students may live in a less favorable family environment with low parental income and education level. This can cause minority students to be less prepared for college than their majority peers. Throughout the application, I will control family backgrounds as confounders when computing the causal gap of minority status.

Prior-Prop 209

Prior Prop 209, racial preference was legal in colleges’ admission selections. Table 2.5 shows the median-unbiased estimators of minority gaps for $k = 5, \cdots, 40$ as well as the propensity score weighted average gap (ATE). I find that the gaps at the lower tail are almost all negative and statistically significant, except for students with science major in UC Santa Cruz and students with non-science major in UC San Diego. It indicates that prior Prop 209, almost all UC campuses implemented racial preferences during admission. In addition, the gaps at the tail are larger for higher ranked campuses such as Berkeley and Los Angeles than the gaps in the rest of the campuses. This supports the conclusion in Arcidiacono et al. (2015) that less-prepared students may have higher graduation probabilities at less-selective schools because they are less behind than their peers.

\textsuperscript{11} As described in Arcidiacono et al. (2015), the raw preparation score ($Y_{i}^{raw}$) for student $i$ is a weighted average of student’s high school GPA ($GPA_{i}$) and their combined verbal and math SAT score ($SAT_{i}$): $Y_{i}^{raw} = \frac{3}{8} \cdot SAT_{i} + 400 \cdot GPA_{i}$. The CPI $Y_{i}$ is the standardized version of $Y_{i}^{raw}$ such that it has mean 0 and standard deviation 1 for the pool of applications to one or more of the UC campuses.
Post-Prop 209

Starting with the entering class of 1998, the UC campuses were subject to a ban on the use of race-based preferences (affirmative action) in admission enacted under Prop 209. In the following, using data of students admitted between 1998 and 2000, I compute the casual effect of minority status on low-scoring students’ CPI across UC campuses and exit major.

Table 2.6 shows that although the average CPI gaps for all campuses remain significant post-Prop 209, for most of the campuses, the tail gaps are not. The significance in average gap does not necessarily reflect the magnitude of racial preference in admission. Tail gap on the other hand, shows that, for the majority of
camperns-major cases, the racial preference post-Prop 209 is insignificant.

Compare Table 2.5 and 2.6, I find heterogeneous responses of UC campuses to Prop 209. The racial gaps in UC Berkeley and UCLA for students with science major and in UC Berkeley, UC Santa Cruz, and UC Riverside for students with non-science major, remain significant after Prop 209. For UC Santa Cruz science major and UC Santa Cruze and Riverside non-science major, the gap became significant post-Prop 209. This is consistent with the finding in Antonovics and Backes (2014) that some campuses (especially Berkeley) responded to the ban of the race-based affirmative action by lowering weights on academic credentials such as SAT scores and increasing weights on family backgrounds in determining admissions. For UC San Diego non-science, the gap main insignificant. For the rest of cases, the gap was significant pre-Prop 209 and became insignificant post-Prop 209 which is in line with the purpose of Prop 209.
I address them simultaneously by relying on the unconfoundedness assumption and extremal quantile asymptotics.

To summarize, through this empirical application, I find that UC campuses modify their admission selection criteria in response to Prop-209. The modifications are heterogeneous across campuses. Post-Prop 209, the minority gaps for some campuses such as San Diego decrease, while for other campuses such as Berkeley, they increase.

2.8 Conclusion

This paper establishes asymptotic theory and inference procedures for an estimator of the unconditional QTE when the quantile index is close or equal to zero. There are two main difficulties: missing data and data sparsity. I address them simultaneously by relying on the unconfoundedness assumption and extremal quantile asymptotics.
When the quantile index is close or equal to zero, I derive a new asymptotic approximation of the finite sample estimator of the QTE and show that standard bootstrap inference is inconsistent. Based on my new asymptotic theory, I propose a new way to construct a uniformly consistent confidence band for extreme QTEs. Last, by using a linear combination of extreme QTE estimators, I propose a median-unbiased estimator and consistent CI for the 0-th QTE.

I then apply the established inference method to estimate the effect of maternal smoking of African American mothers for the lower tail of infants’ birth weights. Although I cannot reject that maternal smoking has no effect on the lower tail of birth weights at the 90% confidence level, I find that the standard bootstrap CI is two times narrower than the new resampling CI developed in this paper. This suggests that the standard bootstrap CI potentially over-rejects.

In my application of estimating the effect of maternal smoking on extremely low birth weights, the existence of other confounders that can jeopardize the selection on observables is possible. For example, mothers who smoke during pregnancy are more likely to adopt other behaviors (drinking, poor nutritional intake, etc.) that could have a negative impact on birth weight. The selection on unobservables can be addressed by using the large cigarette taxation change as an instrumental variable for maternal smoking. Extending the current theory to incorporate IV and conduct inference for the extremal QTE for compliers would be a useful research direction.

In addition, the current robust CI relies on two hard thresholds $\tau_{n,1}$ and $\tau_{n,2}$. When the quantile of interest is between the two thresholds, the robust CI is conservative so that the power property is not optimal. A useful research direction would be to find a unified inference method that works for all extreme, intermediate, and regular order quantiles, and that has better power properties.
3.1 Introduction

Endogenous selection has been recognized as one of the key methodological issues arising in the analysis of microeconomic data since the seminal articles of Gronau (1974) and Heckman (1974). The most common strategy to deal with selection is to rely on instruments that determine selection but not the potential outcome (see, among others, Heckman, 1974, 1979, 1990, (Ahn and Powell, 1993), (Donald, 1995), (Buchinsky, 1998), (Chen and Khan, 2003), (Das et al., 2003), (Newey, 2009) and (Vella, 1998) for a survey). However, in practice, valid instruments are generally difficult to find. Identification at infinity has been proposed in the literature as an alternative solution to the endogenous selection problem, in situations where one is primarily interested in estimating the effects of some covariates on a potential outcome. In particular, Chamberlain (1986) showed that if some individuals face an arbitrarily large probability of selection and the outcome equation is linear, then one can use these individuals to identify the effects of the covariates on the outcome of
interest. Lewbel (2007) generalized this result by proving that identification can be achieved in the context of moment equality models, provided that a special regressor has a support which includes that of the error term from the selection equation (see Lewbel, 2014, for an overview of the special regressor method). Again, in many applications, such a regressor is hard to come by. In a recent article, D’Haultfoeuille and Maurel (2013) have shown that identification in the absence of an instrument is in fact possible without such a covariate. The starting intuition is that, if selection is endogenous, then one can expect the effect of the outcome on selection to dominate those of the covariates for sufficiently large values of the outcome. Following this idea, one can prove identification under the key condition that selection becomes independent of the covariates at infinity, i.e., when the outcome takes arbitrarily large values.

This paper builds on this insight and develops a novel inference method for a class of semiparametric location-scale models subject to endogenous selection. Unlike prior estimation methods for sample selection models, we propose a distribution-free estimator that does not require an instrument for selection nor a large support regressor. Besides, we do not restrict the selection process, apart from the independence at infinity condition mentioned above. We interpret this condition in the context of standard selection models, and show that it translates into a restriction on the copula between the error terms of the outcome and selection equation. This restriction is mild provided that selection is endogenous, and holds for several classical families of copulas, including Gaussian copulas with positive dependence. We also rely on the location-scale specification for identification. This structure, which is reasonable in many settings, imposes that the different subpopulations defined by the covariates have the same distribution of potential outcomes, up to location and scale. The location-scale specification is common in the econometrics literature. In particular,
similar conditions are often imposed in the context of sample selection models with instruments (see, e.g., (Ahn and Powell, 1993), (Buchinsky, 1998), (Chen and Khan, 2003) or (Newey, 2009)) and in the related context of censored regression ((Chen et al., 2005)). Under the independence at infinity assumption, the location and scale functions are identified nonparametrically, using the upper tail of the conditional distribution of the observed outcome. We then show that the covariates effects on the rest of the distribution are generally partially identified, and point identified if the covariate under consideration does not shift the scale.

Turning to estimation, we establish that linear quantile regressions, for large values of the quantile indices, allow us to recover some linear combinations of the covariates effects on the location and scale of the outcome. Those parameters can then be estimated in a second step by a simple minimum distance estimator, which combines the previous estimators for a range of quantile indices. This insight is important for at least two reasons. First, our estimator is simple to implement. In particular, unlike most of the existing semiparametric estimators for sample selection models, our estimator is not based on a nonparametric first step. Second, the asymptotic properties of extremal quantile regressions, that is quantile regressions applied to the tails, have been thoroughly studied in the case without selection in an important paper by Chernozhukov (2005b). This provides a very natural starting point to develop asymptotic inference in our setting. The estimators of the location and scale parameters can then be used to construct bounds on the quantile effects.

We characterize the sharp bounds and further derive simpler outer bounds on which

---

1 Formally, denoting by $n$ the sample size and $\tau_n$ the quantile index, extremal quantile regressions correspond to $\tau_n$-quantile regressions where $\tau_n$ tends to zero as the sample size $n$ grows to infinity. In this paper, we focus on the intermediate order case, which corresponds to situations where $\tau_n \times n$ tends to infinity. See Chernozhukov and Du (2008) for a review of extremal quantile regressions. See also related work by Altonji et al. (2008), who derive the asymptotic properties of a nonparametric extremal quantile regression estimator. While their framework is very general, it cannot be readily extended to the case where the outcome is subject to sample selection.
one can conduct inference using the methodology developed by Chernozhukov et al. (2013b). It is worth noting that, while we use quantile regressions as a tool to circumvent the selection issue, we assume a linear location-scale specification for the potential outcome. This is different in spirit from the methods recently proposed in the literature to estimate quantile regression models in the presence of sample selection (see notably (Arellano and Bonhomme, 2011), and (Melly and Huber, 2011)).

The main difficulty in establishing the asymptotic properties of our estimator is that because of selection, extremal conditional quantiles are not exactly linear here, but only equivalent to a linear form as the quantile index $\tau_n$ tends to zero. Hence, we face a bias-variance trade-off that is typical in non- or semiparametric analysis. Choosing a moderately small quantile index decreases the variance of the estimator, but this comes at the price of a higher bias. Conversely, choosing a very small quantile index mitigates the bias, but increases the variance. In the paper, we provide sufficient conditions under which both bias and variance vanish asymptotically, resulting in asymptotically normal and unbiased estimators. As in the case without selection examined by Chernozhukov (2005b), the convergence rates are not standard, and depend on the tail behavior of the error term from the outcome equation. This is broadly similar to the convergence rates discussed in Andrews and Schafgans (1998), Schafgans and Zinde-Walsh (2002) and Khan and Tamer (2010), the main difference being that the tail behavior of the outcome is going to play a key role here, rather than that of the covariates. Importantly, though, our asymptotic results suggest a rate-adaptive approach for inference, as in Khan and Tamer (2010) and Chernozhukov and Fernandez-Val (2011).

Asymptotic normality and unbiasedness of our estimators requires an appropriate choice of the quantile index, similarly to nonparametric kernel regressions that require an appropriate bandwidth choice. But contrary to the latter case, admissible rates
of convergence towards zero for the quantile index depend in a complicated way on the data-generating process. An analogous issue arises in the estimation at infinity of the intercept of sample selection models (see (Andrews and Schafgans, 1998) and (Schafgans and Zinde-Walsh, 2002)), as well as in the estimation of extreme value indices (see (Drees and Kaufmann, 1998) and (Danielsson et al., 2001)). This is a difficult problem. In the paper, we propose a heuristic data-driven procedure that selects the quantile index minimizing a criterion function capturing the trade-off between bias and variance. In particular, we use subsampling combined with a minimum distance estimator to proxy the bias term, which, in this setting, cannot be simply estimated. Monte Carlo simulation results show that our estimators perform well in finite samples. We further provide evidence that bootstrap, which is shown to be consistent, yields good coverage even in small samples.

Finally, we apply our method to the estimation of the black-white wage gap among males from the 1979 and 1997 cohorts of the National Longitudinal Survey of Youth (NLSY79 and NLSY97). Following Neal and Johnson (1996), we focus on the residual portion of the wage gap that remains after controlling for premarket factors. To the extent that black males are more likely to dropout from the labor market than white males, as was first pointed out in the influential work of Butler and Heckman (1977), correcting for selection is crucial for consistently estimating the black-white differential in terms of potential wages. Besides, evidence that the black-white employment gap has substantially widened over time (see, e.g., (Juhn, 2003) and (Neal and Rick, 2014)) stresses the importance of dealing with selection in order to draw valid conclusions regarding the across-cohort evolution of the black-white wage gap. In this context, finding a valid instrument that affects selection but not potential wages is particularly challenging, making it desirable to use an estimation method that does not require such an instrument. For the NLSY79 cohort, we find
a smaller residual wage gap (10.1%) than the one obtained using the imputation method of Neal and Johnson (1996) and Johnson et al. (2000), which is consistent with our approach being based on a weaker identifying restriction. Overall, our estimates strengthen the key takeaway of Neal and Johnson (1996) by providing evidence of an even more important role played by the black-white AFQT gap.

Turning to the evolution across the 1979 and 1997 cohorts, we find that there has been a slow convergence in the raw male black-white wage gap between 1990 and 2007 (-4.6 pp), and an even slower convergence in the residual portion of the wage gap that remains after controlling for premarket factors such as AFQT and family background (-1.2 pp). Interestingly, this provides evidence that premarket skills are a key component of the level as well as the evolution of the black-white wage gap. Besides, that the wage gap remaining after accounting for differences in premarket factors is essentially stable after almost 20 years suggests that this residual portion of the wage gap is an important factor behind the slow convergence of the wages of blacks and whites.

The remainder of the paper is organized as follows. Section 2 presents the set-up and discusses the identification results. Section 3 defines the estimators and establishes the main asymptotic normality results. Section 4 discusses some Monte Carlo simulation results. Section 5 applies our method to the estimation of the black-white wage gap among males. Finally, Section 6 concludes. Additional details on the estimation procedure and the data, along with the proofs, are collected in the appendix.

2 Other noteworthy papers analyzing the black-white wage gap while using imputation methods to correct for selection into the workforce include Brown (1984), Smith and Welch (1989), Juhn (2003), Neal (2004), Neal (2006) and Neal and Rick (2014).
3.2 The set-up and nonparametric identification

3.2.1 Model and main result

Before presenting the model, let us introduce some notations and definitions. For any random variable $U$, we denote by $F_U$ and $S_U$ its cumulative distribution function (cdf.) and survival function, while $Q_U$ denotes its quantile function, $Q_U(u) = \inf\{u : F_U(u) \geq \tau\}$. For more general increasing functions $G$, we let $G^{-1}(u) = \inf\{v : G(v) \geq u\}$, with the convention that $\inf \emptyset = +\infty$, denote its generalized inverse. Finally, we use in the following some notions from extreme value theory. A function $F$ is regularly varying at $x \in (0, +\infty)$ with index $\alpha \in (-\infty, +\infty]$, and we write $F \in RV_{\alpha}(x)$, if for any $t > 0$, $\lim_{u \to x} F(tu)/F(u) = t^\alpha$, with the understanding that $t^\infty = \infty$ if $t > 1$ and $= 0$ if $1 > t > 0$ (and similarly for $\alpha = -\infty$). $F$ is slowly varying at $x$ if $F \in RV_0(x)$. We also say that a given cdf. $F$ belongs to the domain of attraction of generalized extreme value distributions if there exists sequences $(a_n)_{n \in \mathbb{N}}$ and $(b_n)_{n \in \mathbb{N}}$ and a cdf. $G$ such that for any independent draws $(U_1, \ldots, U_n)$ from $F$, $b_n^{-1}(\max(U_1, \ldots, U_n) - a_n)$ converges in distribution to $G$. In such a case, $G$ belongs to the family of generalized extreme value distributions. Finally, we denote by $X_j$ the $j$th component of any given random vector $X \in \mathbb{R}^d$, and for any function $\phi$ differentiable with respect to its $j$th argument, we denote the partial derivative of $\phi$ with respect to that argument by $\partial_j \phi$.

Let $Y^*$ denote the outcome of interest and $X \in \mathbb{R}^d$ denote a vector of covariates, excluding the constant. We suppose that $Y^*$ and $X$ are related through the nonparametric location-scale model

$$Y^* = \psi(X) + \sigma(X) \varepsilon,$$  \hspace{1cm} (3.2.1)

where we suppose, without loss of generality, that $\sigma(X) > 0$. Because we do not standardize $\varepsilon$, we can always suppose that there is a $x_0 \in \text{Supp}(X)$ such that $\psi(x_0) =$
0 and $\sigma(x_0) = 1$. Our focus throughout the paper is on $\psi(\cdot)$, $\sigma(\cdot)$, along with the quantile effects of $X_j$ on $Y^*$. These quantile effects correspond to the effect on $Y^*$ of an exogenous, infinitesimal, change of $X_j$, or a change from $X_j = 0$ to $X_j = 1$ if $X_j$ is binary, for individuals at a given conditional quantile of $Y^*$.

We face a sample selection issue here as we only observe $(D, Y = DY^*, X)$, where $D$ denotes the selection dummy. Importantly, we do not assume to have access to an instrument affecting $D$ but not $Y^*$, nor do we require one of the covariates to have a large support. Instead, identification is achieved under the following conditions.

**Assumption 19.** *(Exogeneity)* $X \perp \varepsilon$.

**Assumption 20.** *(Tail and regularity of the residual)* (i) $\sup(\text{Supp}(\varepsilon)) = \infty$, (ii) $S_{\exp(\varepsilon)}$ is not slowly varying at infinity, (iii) $S_\varepsilon$ is in the domain of attraction of generalized extreme value distributions and (iv) the distribution of $(X, \varepsilon)$ conditional on $D = 1$ is dominated by a product measure. We denote by $f_{\varepsilon|D=1,X}$ and $f_{Y|D=1,X}$ the corresponding conditional densities.

**Assumption 21.** *(Independence at infinity)* There exists $h \in (0, 1]$ such that for all $x \in \text{Supp}(X)$,

$$\lim_{y \to \infty} P(D = 1|X = x, Y^* = y) = h.$$

Assumption 19 is restrictive but commonly made in the context of selection models. It is also weaker than the exogeneity assumption imposed, for instance, by Chamberlain (1986) or Ahn and Powell (1993), since we allow for heteroskedasticity here. Assumption 20-(ii) is satisfied if, for instance, $E(\exp(b\varepsilon)) < \infty$ for some $b > 0$.

---

3 Notable exceptions include Das et al. (2003) and Lewbel (2007), who allow for endogenous regressors. However, the estimators proposed in these papers require an instrument for selection or a special regressor, respectively.
Note that this tail condition is fairly mild. For example, in the context of a wage equation where $Y^*$ corresponds to the logarithm of the wage $W$, it is satisfied as long as $E(W^b) < \infty$ for a given $b > 0$. It follows that this condition holds even if wages exhibit very fat tails, for instance Pareto-like. Conditions (i) and (iii) are not necessary for identification but will be used subsequently. Condition (iii) is mild and satisfied by most of the standard continuous cdf., including the normal one. Condition (iv), which is very mild, is not needed for the identification of $\psi(.)$ and $\sigma(.)$. It is only required to define the bounds on the quantile effects.

Finally, Assumption 21 is our main identifying condition. The key part is that $h$ does not depend on $x$. In other words, we require selection to become independent of the covariates at infinity, that is conditional on having arbitrarily large outcomes. The underlying intuition is that, if selection is endogenous, then one can expect the effect of the outcome on selection to dominate those of the covariates for sufficiently large values of the outcome. This condition includes as an important special case the “no selection at infinity” situation where the selection probability tends to one for large values of the outcome ($h = 1$). But our framework also accommodates more general forms of selection since $h < 1$ is also allowed for. In the context of labor market participation, Assumption 21 holds as long as the set of individuals with arbitrarily large potential wages join the workforce with a positive constant probability $h$. In practice, one can imagine that $h$ would be smaller than one if some of those high-productivity individuals were not able to enter the labor market for unobserved idiosyncratic reasons, such as adverse health shocks. We return to this condition in the following section, by discussing in detail several examples where this condition holds.

The first part of Theorem 3.2.1 below states that, under Assumptions 19-21, the location and scale functions $\psi(.)$ and $\sigma(.)$ are identified. The second part of the
Theorem provides bounds on the quantile effects, which are sharp under the following regularity conditions.

Assumption 22. (Regularity conditions for the sharpness of the bounds) (i) $\text{Supp}(X)$ is compact, (ii) $\inf_{x \in \text{Supp}(X)} \psi(x) > -\infty$ and $\inf_{x \in \text{Supp}(X)} \sigma(x) > 0$ and (iii) there exists $K > 0$ such that for all $(x, x') \in \text{Supp}(X)^2$ and all $y$ large enough,

$$|P(D = 1|X = x, Y^* = y) - P(D = 1|X = x', Y^* = y)| \leq K \|x - x'\|, \quad (3.2.2)$$

where $\|x\|$ denotes the euclidian norm of $x$.

Theorem 3.2.1. Under Assumptions 19-21, $\psi(.)$ and $\sigma(.)$ are identified. Suppose also that $\psi(.)$ and $\sigma(.)$ are differentiable with respect to their $j$th argument. Then the quantile effect $\Delta_{j\tau}(x)$ satisfies

$$\Delta_{j\tau}(x) = \partial_j \psi(x) + \partial_j \sigma(x) \int_{-\infty}^{\tau} [\tau - 1\{\partial_j \sigma(x) > 0\}(1 - F_x(+\infty))] \, du,$$

$$\Delta_{j\tau}(x) = \partial_j \psi(x) + \partial_j \sigma(x) \int_{-\infty}^{\tau} [\tau - 1\{\partial_j \sigma(x) < 0\}(1 - F_x(+\infty))] \, du,$$

and where

$$F_{\varepsilon}(v) = \int_{-\infty}^{\tau} \sup_{x \in \text{Supp}(X)} \left[ P(D = 1|X = x) \sigma(x) f_{Y|D=1,X}(\psi(x) + \sigma(x) u|x) \right] du.$$

Finally, suppose that Assumption 22 also holds. Then the bounds $\underline{\Delta}_{j\tau}(x)$ and $\overline{\Delta}_{j\tau}(x)$ are sharp.

The underlying intuition of the identification result for $\psi(.)$ and $\sigma(.)$ is that under Assumption 21, the right tail of $Y$ and $Y^*$ are equivalent up to a multiplicative constant. It follows that we can use the conditional survival function of $Y$ given $X$ to uniquely recover the location and scale functions, provided the tail of the residual

\[\text{when } X_j \text{ is binary, } \Delta_{j\tau} \text{ should rather be defined by } \Delta_{j\tau} = \text{Q}_{Y*|X_{-j},X_j=1} - \text{Q}_{Y*|X_{-j},X_j=0},\]

where $X_{-j}$ denotes all components of $X$ except $X_j$. We obtain the same formulas for the bounds, except that partial derivatives should be replaced by differences.
is not too fat (Assumption 20-(ii)). However, point identification of \( \psi(.) \) and \( \sigma(.) \) does not necessarily entail point identification of the quantile effects. This is due to the fact that \( \Delta_{j\tau}(x) = \partial_j\psi(x) + \partial_j\sigma(x)Q_\varepsilon(\tau) \). Because of the missing data issue, the quantile \( Q_\varepsilon(\tau) \) cannot be point identified in general. Specifically, \( \Delta_{j\tau}(x) \) is point identified and equal to \( \partial_j\psi(x) \) if \( \partial_j\sigma(x) = 0 \). \( \Delta_{j\tau}(x) \) is only partially identified otherwise. Note that the condition \( \partial_j\sigma(x) = 0 \) is much weaker than assuming that the residuals are homoskedastic with respect to \( X \), as it only requires that the \( j \)th component of \( X \) does not shift the scale of the outcome.

An important limitation of the sharp bounds is that, to the best of our knowledge, no existing method can be readily used to conduct inference on them. Kitagawa (2010) provides some useful results, but in a simpler framework where \( X \) is discrete and without the need to estimate \( \psi(.) \) and \( \sigma(.) \) in a first step. On the other hand, we derive in Appendix B.3.2 the following outer bounds on \( \Delta_{j\tau}(x) \):

\[
\Delta^o_{j\tau}(x) = \partial_j\psi(x) + \partial_j\sigma(x) \left[ 1 \{ \partial_j\sigma(x) > 0 \} Q^o_\varepsilon(\tau) + 1 \{ \partial_j\sigma(x) < 0 \} \overline{Q}^o_\varepsilon(\tau) \right]
\]

\[
\overline{\Delta}^o_{j\tau}(x) = \partial_j\psi(x) + \partial_j\sigma(x) \left[ 1 \{ \partial_j\sigma(x) > 0 \} \overline{Q}^o_\varepsilon(\tau) + 1 \{ \partial_j\sigma(x) < 0 \} Q^o_\varepsilon(\tau) \right]
\]

with

\[
Q^o_\varepsilon(\tau) = \sup_{x \in \text{Supp}(X)} Q_{Y|D=1,X=x} \left( \frac{\tau - P(D=0|X=x)}{P(D=1|X=x)} \right) - \psi(x) \]

\[
\overline{Q}^o_\varepsilon(\tau) = \inf_{x \in \text{Supp}(X)} Q_{Y|D=1,X=x} \left( \frac{P(D=1|X=x)}{\tau} \right) - \psi(x) \]

(3.2.3)

The outer bounds take a more convenient form for inference than the sharp bounds. Besides, they actually coincide with these sharp bounds when the family of functions \( e \mapsto P(D = 1|X = x)\sigma(X)f_{Y|D=1,X}(\psi(X) + \sigma(X)e|x) \) indexed by \( x \) do not intersect.
3.2.2 The independence at infinity condition

Point identification of $\psi(.)$ and $\sigma(.)$ relies mostly on the independence at infinity assumption. To get a better sense of this condition, we discuss it below in the context of two common selection models. The first one is a threshold crossing model described in Assumption 23.\(^5\)

**Assumption 23.** (i) $D = 1\{\phi(X) - \eta \geq 0\}$ with $(\varepsilon, \eta) \perp\!\!\!\!\perp X$, (ii) $\inf_{x \in \text{Supp}(X)} F_{\eta}(\phi(x)) = \upsilon > 0$, (iii) $F_{\varepsilon}$ and $F_{\eta}$ are continuous and strictly increasing and the copula $C$ of $(-\varepsilon, \eta)$ is differentiable with respect to its first argument.

The first condition defines the selection model as a standard threshold crossing model. Importantly however, we do not add any instrument in this selection equation. The second condition ensures that $x \mapsto P(D = 1|X = x)$ is bounded below by a positive number. Note that this condition will typically hold if none of the covariates has a large support, which is precisely the type of situation we are interested in. In this context, Proposition 3.2.1 provides a restriction on $C$ ensuring that Assumption 21 is satisfied. Hereafter, let $f_C(\tau) = \sup_{u \leq \tau, v \in [0,1]} |\partial_1 C(u, v) - 1|$.

**Proposition 3.2.1.** Suppose that Assumptions 19, 20 and 23 hold, and

$$
\lim_{\tau \to 0} f_C(\tau) = 0. \tag{3.2.4}
$$

Then Assumption 21 is satisfied, and therefore $\psi(.)$ and $\sigma(.)$ are identified.

The key idea is that selection becomes independent of the covariates for large values of the outcome if selection is endogenous enough, in the sense that $(-\varepsilon, \eta)$ satisfies (3.2.4). To understand this condition better, it is useful to consider two extreme cases. In the perfect dependence case such that $\eta = -\varepsilon$, then $\partial_1 C(u, v) = 1$

---

\(^5\) Our identification strategy is also natural in the context of the generalized Roy model (Heckman and Vytlacil, 2007). We refer the reader to D’Haultfoeuille et al. (2014) for a detailed discussion on this question.
for all \( u < v \), so that (3.2.4) actually holds exactly for small values of \( \tau \). On the other hand, when \( \eta \) and \( -\varepsilon \) are independent, \( \partial C(u, v) = v \), and \( f_C(\tau) = 1 - v \), which is positive except in the degenerate case where \( D = 1 \) almost surely. In between these two extreme cases, Table 3.1 provides examples of copulas that satisfy this constraint. It underlines that Assumption 21 may be satisfied even if the dependence between \( -\varepsilon \) and \( \eta \) is very weak. Importantly, it holds for all Gaussian copulas with positive dependence.\(^6\) It also holds for Archimedean copulas under a restriction on the behavior of the generator \( \Psi \) around 0. This restriction holds for instance for the Clayton copula, for which \( \Psi(u) = (u^{-\theta} - 1)/\theta \), provided that \( \theta > 0 \). The Gumbel family is another popular Archimedean family of copulas that does not satisfy the restriction on \( \Psi \), since \( \Psi \) is slowly varying at 0. However, Condition (3.2.4) still holds for some parameters of this family.

Table 3.1: Examples of copulas satisfying (3.2.4).

<table>
<thead>
<tr>
<th>Copula family</th>
<th>Restriction ensuring (3.2.4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian ( C(u, v; \rho) )</td>
<td>( \rho &gt; 0 )</td>
</tr>
<tr>
<td>Archimedean ( C(u, v; \Psi) = \Psi^{-1}(\Psi(u) + \Psi(v)) )</td>
<td>( \lim_{u \to 0} \Psi(u) = +\infty ) ( \Psi ) is ( C^1 ) and ( RV_\alpha(0) ) with ( \alpha \in (0, +\infty] )</td>
</tr>
<tr>
<td>Gumbel ( \Psi(u; \theta) = (-\log(u))^\theta )</td>
<td>( \theta &gt; 1 )</td>
</tr>
</tbody>
</table>

3.3 Semiparametric estimation

3.3.1 Definition of the estimators

In this section, we consider the estimation of a semiparametric version of Model (3.2.1), by imposing the following condition. We start by defining our estimators before establishing their asymptotic properties in the next subsection. Hereafter, we

\(^6\) Note that what is important here is the strength, but not the sign, of the dependence between \( \eta \) and \( -\varepsilon \). The case of negative selection could be addressed by replacing \( Y^* \) by \( -Y^* \) and \( \varepsilon \) by \( -\varepsilon \).
denote by $X$ the vector $[1, X']'$.  

**Assumption 24.** (Covariates) $\psi(X) = X'\beta$ and $\sigma(X) = 1 + X'\delta$. Moreover, $X$ has a compact support $\text{Supp}(X)$ and $Q_X = E(XX')$ is full rank.

The linearity assumption will allow us to use standard quantile regressions, thus leading to computationally simple estimators. This specification also results in faster convergence rates relative to a more general nonparametric location-scale specification. Note that we do not include the intercept in $X'\beta$ and $X'\delta$, and fix the constant in $\sigma(X)$ to one. These location and scale normalizations are innocuous since we do not standardize $\varepsilon$.

Suppose we have a sample $(D_i, Y_i, X_i)_{i=1...n}$ of $n$ i.i.d. random variables distributed as $(D, Y, X)$. The starting point for estimation is that under Assumptions 19, 21 and 24, we have, as $y \to -\infty$,

$$F_{-Y|X}(y|x) \sim h F_{-\varepsilon}((y + x'\beta)/(1 + x'\delta)) \quad (3.3.1)$$

The key insight for estimating $(\beta, \delta)$ is that if one also imposes Assumption 20, then it is possible to invert both sides and maintain the equivalence. It follows that the quantile regression of $-Y$ on $X$ is asymptotically linear. This result is going to play an important role in our estimation procedure.

**Lemma 3.3.1.** Under Assumptions 19-21 and 24, as $\tau \to 0$,

$$Q_{-Y|X}(\tau|x) \sim \gamma(\tau) + x'\beta(\tau) \quad (3.3.2)$$

where $\gamma(\tau) = Q_{-\varepsilon}(\tau/h)$ and $\beta(\tau) = -\beta + \gamma(\tau)\delta$.

Lemma 3.3.1 provides the intuition that it might be possible to use quantile regressions in the tails to consistently estimate $(\gamma(\tau), \beta(\tau))$, for small values of the quantile index $\tau$. The main difficulty in formalizing this intuition though, comes
from the fact that (3.3.2) is an equivalence and not an equality, which gives rise to a bias term that needs to be controlled. We define

$$(\hat{\gamma}(\tau), \hat{\beta}(\tau)) = \arg\min_{\gamma, \beta} \sum_{i=1}^{n} \rho_{\tau}(-Y_i - \gamma - X_i'\beta),$$

where $\rho_{\tau}(u) = (\tau - 1\{u < 0\})u$ is the check function used in quantile regressions.

Then one can simply use the following relationships to estimate the parameters of interest $\beta$ and $\delta$:

$$\delta = \frac{\beta(l_1\tau) - \beta(\tau)}{\gamma(l_1\tau) - \gamma(\tau)},$$

$$\beta = -\beta(\tau) + \gamma(\tau)\delta,$$

where the first equality holds provided that $\gamma(l_1\tau) - \gamma(\tau) \neq 0$. We basically follow this route in the paper, except that for an efficiency matter we estimate $\delta$ using $J$ reduced form estimators $(\beta(l_1\tau), \ldots, \beta(l_J\tau))$ rather than just two, where $(l_1, \ldots, l_J)$ is a vector of positive spacing parameters such that $l_j \neq 1$ for all $j \in \{1, \ldots, J\}$. Let us consider

$$g_n(\delta) = \left( \frac{\hat{\beta}(l_1\tau) - \hat{\beta}(\tau) - (\hat{\gamma}(l_1\tau) - \hat{\gamma}(\tau))\delta}{\hat{\gamma}(l_1\tau) - \hat{\gamma}(\tau)} \right)$$

and let $W_n$ be a $Jd \times Jd$ positive definite symmetric matrix. We estimate $\delta$ using a minimum distance procedure:

$$\hat{\delta} = \arg\min_{\delta} g_n(\delta)'W_n g_n(\delta). \quad (3.3.3)$$

Finally, we estimate $\beta$ by averaging across the quantile indices:

$$\hat{\beta} = \frac{1}{J+1} \sum_{j=0}^{J} -\hat{\beta}(l_j\tau) + \hat{\gamma}(l_j\tau)\hat{\delta},$$

Assumptions 20-(i) and 26 below ensure that the latter condition is satisfied for any $l \neq 1$ and $\tau$ small enough.
with $l_0 = 1$. We do not simultaneously estimate $\beta$ and $\delta$ since the corresponding estimators of $\beta$ and $\delta$ would have different rates of convergence, thus implying that the standard asymptotic theory of minimum distance estimators would not apply in this context. In particular, this framework would lead to a singular optimal weighting matrix. Intuitively, only the terms with the slowest rate of convergence would be weighted positively, since the other terms would not matter asymptotically. We would then lose consistency of the estimators.

Our estimators depend on a choice of $\tau$, $(l_1, ..., l_J)$ and $W_n$. We derive in the following section the optimal weighting matrix, which can be consistently estimated. Regarding the quantile indices, while the choice of the constants $(l_1, ..., l_J)$ does not appear to matter much in practice, an appropriate choice of $\tau$ is crucial to balance bias and variance in such a way that guarantees our estimators to be asymptotically normal with zero mean. We propose a data-driven procedure for that purpose in Section 3.3.3.

We now turn to the estimation of the quantile effects, which in this semiparametric setting are given by $\Delta_{j\tau} = \beta_j + \delta_j Q_x(\tau)$. $\Delta_{j\tau}$ is point identified when $\delta_j = 0$ and partially identified otherwise. Moreover, as shown in Section 3.3.2, it is possible to estimate $\beta_j (= \Delta_{j\tau})$ more precisely than with the previous unconstrained estimator $\hat{\beta}_j$ in the case where $\delta_j = 0$. We consider these two cases separately, noting that the restriction $\delta_j = 0$ can be tested using the asymptotic distribution of $\hat{\delta}_j$ provided in the following section.

Suppose first that the model is partially homoskedastic, in the sense that $\{\delta_{jk}\}_{k=1}^{d_{\beta}}$ are equal to zero for some $d \geq d_{\beta} \geq 1$ and $\{j_1, ..., j_{d_{\beta}}\} \subset \{1, ..., d\}$. Then the quantile effects of $\{X_{jk}\}_{k=1}^{d_{\beta}}$ are equal to the average effects of the corresponding covariates, which are identified and equal to $\{\beta_{jk}\}_{k=1}^{d_{\beta}}$. Let $\Psi$ be a $d_{\beta} \times d$ matrix that picks out
the corresponding subvector of $\beta$, i.e. $\Delta^1 = \Psi \beta$ and let us consider

$$g_{1n}(\Delta^1) = \left( \Psi \hat{\beta}(\tau) - \Delta^1, \Psi \hat{\beta}(l_1 \tau) - \Delta^1, ..., \Psi \hat{\beta}(l_J \tau) - \Delta^1 \right)'.$$

We then propose to estimate $\Delta^1$ by

$$\hat{\Delta} = \arg \min_{\Delta} g_{1n}(\Delta)' W_{1n} g_{1n}(\Delta),$$

for some positive definite matrix $W_{1n}$.

Finally, for the components $j$ such that $\delta_j \neq 0$ so that the quantile effects are only partially identified, one possibility would be to estimate the sharp bounds given in Theorem 3.2.1. As mentioned previously, however, to the best of our knowledge one cannot conduct inference on these bounds using available methods. Instead, we propose to use a semiparametric version of the simpler outer bounds given by (3.2.3). We focus here on the upper bound $\overline{\Delta}_j \tau$ on $\Delta_j \tau$, when $\delta_j > 0$. The lower bound and the case $\delta_j < 0$ are obtained similarly. Let

$$\theta_j(x) = \beta_j + \delta_j \frac{Q_{Y|D=1,X=x}(\overline{P}(D=1|X=x)) - x' \beta}{1 + x' \delta},$$

so that $\overline{\Delta}_j \tau = \inf_{x \in \text{Supp}(X)} \theta_j(x)$. One can estimate $\theta_j(.)$ by

$$\hat{\theta}_j(x) = \hat{\beta}_j + \delta_j \frac{\hat{Q}_{Y|D=1,X=x}(\overline{\hat{P}}(D=1|X=x)) - x' \hat{\beta}}{1 + x' \delta},$$

where $\hat{Q}_{Y|D=1,X=x}$ and $\overline{\hat{P}}(D = 1|X = x)$ are nonparametric (for instance kernel) estimators of $Q_{Y|D=1,X=x}$ and $P(D = 1|X = x)$, respectively. Then the bounds can be consistently estimated using the plug-in estimators $\hat{\Delta}^\omega_{j \tau} = \inf_{x \in \text{Supp}(X)} \hat{\theta}_j(x)$, where $\text{Supp}(X)$ denotes the observed values of $X$ in the sample.
3.3.2 Asymptotic properties and inference

Location and scale parameters

We first consider the asymptotic properties of \((\hat{\beta}, \hat{\delta})\). We rely for that purpose on the asymptotic properties of extremal quantile regressions, established by Chernozhukov (2005b). As already discussed, an important difference is that (3.3.2) is an equivalence rather than an equality. This implies that a bias term comes into play, which needs to be controlled.

In addition to Assumptions 19-21 and 24, our asymptotic analysis relies on the three conditions discussed below. In the following, we let

\[ f(\gamma) = E \left[ \sup_{u \leq \gamma} |h - P(D = 1|X, -\varepsilon = u)| \times |X\| \right]. \]

**Assumption 25.** (i.i.d. sampling) \((D_i, Y_i, X_i)_{i=1,...,n}\) are independent, with the same distribution as \((D, Y, X)\).

**Assumption 26.** (Monotone densities) There exists \(A_0\) such that almost surely, \(F_{-\varepsilon}\) and \(F_{-\varepsilon|D=1,X}\) are differentiable with increasing derivatives on \((-\infty, A)\).

**Assumption 27.** (Rate of convergence of the quantile index) \(\tau_n\) satisfies, as \(n \to \infty\), (i) \(\tau_n \to 0\), (ii) \(\tau_n n \to \infty\) and (iii) \(\sqrt{\tau_n} n f(\gamma(\tau_n)) \to 0\), where \(\gamma(\tau_n) = Q_{-\varepsilon}(\tau_n/h)\).

Assumption 26 rules out erratic behavior of the densities in the tail. It is very mild and satisfied by all standard distributions. Assumption 27 is an important condition that restricts the rate of convergence of the tail index \(\tau_n\). Conditions (i) and (ii) basically ensure that the number of observations that are useful for inference, which is proportional to \(\tau_n n\), tends to infinity, but at a slower rate than the sample size. Thus, following the standard terminology in order statistics theory, our estimators are based on quantile regressions where \(\tau_n\) is an intermediate order sequence, which we will refer to as intermediate order quantile regressions. The reason why we use
intermediate order instead of extreme order sequences, where $\tau_n$ tends to a non-zero constant, is that in the latter case, $\hat{\delta}$, and thus $\hat{\beta}$, are not consistent. Intuitively, this is due to the fact that only a finite number of observations are useful in the extreme order case. Intermediate order quantile theory also has the nice feature that it guarantees asymptotic normality rather than convergence towards a non-standard, data-dependent, distribution (see (Chernozhukov, 2005b) and (Chernozhukov and Fernandez-Val, 2011), in the absence of sample selection). Finally, Condition (iii) is specific to our context. This is an undersmoothing condition, which ensures that the bias arising because (3.3.2) is an equivalence rather than an equality vanishes quickly enough.

Importantly, under Assumption 21, there always exists a $\tau_n$ satisfying Assumption 27. To see this, observe first that if $f = 0$ on $(-\infty, B)$ for some $B$, then (iii) holds trivially. Otherwise define, for any $\alpha \in (0, 1)$, $G(\gamma) = F_{-\varepsilon}(\gamma) f(\gamma)^{2(1-\alpha)}$. By construction, $f$ is increasing. Because $F_{-\varepsilon}$ is strictly increasing on $(-\infty, A)$ by Assumption 26, $G$ is also strictly increasing on $(-\infty, A)$. Then define, for $n$ large enough,

$$\tau_n^* = h F_{-\varepsilon} \circ G^{-1}(1/n).$$

(3.3.4)

Under Assumption 21, $\lim_{\gamma \to -\infty} f(\gamma) = 0$. Thus, $\lim_{\gamma \to -\infty} G(\gamma) = 0$. This implies that $\lim_{n \to \infty} G^{-1}(1/n) = -\infty$, ensuring that $\tau_n^*$ satisfies Condition (i). Moreover, it follows from the equality $F_{-\varepsilon}(\gamma) = \frac{G(\gamma)}{f^{2(1-\alpha)}(\gamma)}$ that

$$\tau_n^* = \frac{h/n}{f^{2(1-\alpha)} \circ G^{-1}(1/n)},$$

which implies that Condition (ii) holds as well. Finally, by using this expression

---

8 To see this, note that for any $x$, $\sup_{u \leq y} |h - P(D = 1|X = x, -\varepsilon = u)|$ tends to zero by Assumption 21. Because this term is bounded by 2, $f$ tends to zero by the dominated convergence theorem.
again and noting that $\gamma(\tau_n^*) = G^{-1}(1/n)$, we get
\[
\sqrt{n\tau_n^*} f(\gamma(\tau_n^*)) = \frac{\sqrt{h \times f \circ G^{-1}(1/n)}}{f(1-a) \circ G^{-1}(1/n)} = \sqrt{h} f^n \circ G^{-1}(1/n),
\]
so that Condition (iii) is also satisfied. An obvious issue is that such a $\tau_n^*$ depends on $F_\varepsilon$ and $f$, both of which are unknown to the researcher. We shall come back to the issue of the practical choice of $\tau_n$ in Section 3.3.3.

The main result of this subsection is stated in Theorem 3.3.1 below, which shows that the estimators of $\beta$ and $\gamma$ are consistent, asymptotically normal, and that the bootstrap is consistent for both. Before stating the result, we need to introduce several matrices. First, let $L$ be the matrix of typical term $L_{i,j} = l_{i,j}^{l_{i,j}}$ for $(i, j) \in \{1, ..., J+1\}^2$. Second, let $\Delta = [-\Delta, I_d]$, where $I_d$ denotes the identity matrix of size $d$. Define $\Gamma = [-\iota, \text{diag}(1/\sqrt{I_1}, ..., 1/\sqrt{I_J})] \otimes I_{d+1}$, where $\iota$ denotes the column vector of ones of size $J$ and, for any vector $v$, $\text{diag}(v)$ denotes the diagonal matrix with diagonal $v$. Finally, let $G = (\log(l_1), ..., \log(l_J))' \otimes I_d$, $Q_H = E\left[XX'/(1 + X'\delta)\right]$ and $\Omega_0 = Q_{H}^{-1}Q_XQ_{H}^{-1}$.

**Theorem 3.3.1.** Under Assumptions 19-21 and 24-27, and if $W_n \xrightarrow{p} W$ symmetric positive definite and nonstochastic,

\[
\sqrt{\tau_n^*} (\hat{\beta} - \beta) \quad \xrightarrow{d} \quad N(0, \Omega_\delta)
\]

\[
\sqrt{\tau_n^*} (\hat{\gamma} - \gamma) \quad \xrightarrow{d} \quad N(0, \Omega_\delta)
\]

where $\Omega_\delta = (G'WG)^{-1}G'W(I_J \otimes \Delta)\Gamma(L \otimes \Omega_0)\Gamma'(I_J \otimes \Delta')WG(G'WG)^{-1}$. The optimal weighting matrix is $W^*_\delta = ((I_J \otimes \Delta)\Gamma(L \otimes \Omega_0)\Gamma'(I_J \otimes \Delta'))^{-1}$ and the corresponding

\footnote{Consider an estimator $\hat{\theta}$ of a parameter $\theta$ such that for some sequence $(r_n)$, $r_n(\hat{\theta} - \theta)$ converges in distribution, and let $\hat{\theta}^*$ denote the bootstrap counterpart of $\hat{\theta}$. We say that the bootstrap is consistent for $\hat{\theta}$ if with probability tending to one and conditional on the sample, $r_n(\hat{\theta}^* - \theta)$ converges to the same distribution as $r_n(\hat{\theta} - \theta)$. We refer to, e.g., van der Vaart and Wellner (1996), Section 3.6, for a formal definition of conditional convergence.}
asymptotic variance is $\Omega_\delta = (G^t W_\delta G)^{-1}$. Moreover, the bootstrap is consistent for both $\hat{\delta}$ and $\hat{\beta}$.

A consistent estimator of the asymptotic variance $\Omega_\delta$ can be obtained by replacing $W$ by $W_n$, $\Delta$ by $[-\hat{\delta}, I_d]$ and $\Omega_0$ by $\hat{\Omega}_H^{-1} \hat{Q}_X \hat{Q}_H^{-1}$, with

$$
\hat{Q}_X = \frac{1}{n} \sum_{i=1}^n X_iX_i', \quad \hat{Q}_H = \frac{1}{n} \sum_{i=1}^n X_iX_i'/(1 + X_i'\hat{\delta}).
$$

Similarly, one can consistently estimate $W_\delta^*$, and thus obtain a two-step estimator that is optimal in the class of estimators considered here.

Theorem 3.3.1 shows that the rates of convergence of $\hat{\delta}$ and $\hat{\beta}$ depend on $\tau_n$, which itself depends on $f(\cdot)$. Intuitively, if $f(\gamma)$ tends quickly to 0 as $\gamma \to -\infty$, $\tau_n$ can actually tend to zero very slowly while still satisfying Assumption 27-(iii), implying fast rates of convergence. Proposition 3.3.1 formalizes this intuition, and provides a condition for consistency of $\hat{\beta}$. Note that this does not directly follow from Theorem 3.3.1, since $|\gamma(\tau_n)| \to \infty$ as $\tau_n \to 0$.

**Proposition 3.3.1.** Suppose that Assumptions 19-21, 24-27 hold and $W_n \xrightarrow{p} W$.

Then there exists $\tau'_n$ satisfying Assumption 27 such that

- $\hat{\beta}$ is consistent if, for some $a > 1$, $f(u) = o(|u|^{-a})$ as $u \to -\infty$.

- The rates of convergence of $\hat{\delta}$ and $\hat{\beta}$ are polynomial if for some $a > 0$, $f(u) = o(F_{-\epsilon}(u)^a)$ as $u \to -\infty$.

Assumption 20 implies that for all $a > 0$, $F_{-\epsilon}(u) = o(|u|^{-a})$. Thus, the condition $f(u) = o(F_{-\epsilon}(u)^a)$ is stronger than the one ensuring consistency of $\hat{\beta}$, as expected. To understand further what both conditions mean, it is useful to discuss them in the context of the sample selection model defined by Assumption 23. In such a case,
letting $\mathfrak{x} = [1, x']$, we have

$$f(\gamma(\tau)) \leq \sup_{\mathfrak{x} \in \text{Supp}(X)} ||\mathfrak{x}|| \sup_{\mathfrak{x} \in \text{Supp}(X), u \leq \gamma(\tau)} |P(D = 1|X = x, -\varepsilon = u) - 1|$$

$$= \sup_{\mathfrak{x} \in \text{Supp}(X)} ||\mathfrak{x}|| \sup_{\mathfrak{x} \in \text{Supp}(X), u \leq \gamma(\tau)} |P(\eta \leq \phi(x)| - \varepsilon = u) - 1|$$

$$= \sup_{\mathfrak{x} \in \text{Supp}(X)} ||\mathfrak{x}|| \sup_{\mathfrak{x} \in \text{Supp}(X), u \leq \gamma(\tau)} |P(F_{\eta}(\eta) \leq F_{\eta}(\phi(x))|F_{-\varepsilon}(-\varepsilon) = F_{-\varepsilon}(u)) - 1|$$

$$\leq \sup_{\mathfrak{x} \in \text{Supp}(X)} ||\mathfrak{x}|| f_C(\tau),$$

where the second inequality follows from $f_C(\tau) = \sup_{u \leq \tau, \alpha \in [2, 1]} |\partial_1 C(u, v) - 1|$. Hence, consistency of $\hat{\beta}$ is achieved if $f_C(\tau) = o(|\gamma(\tau)|^{-a})$ for some $a > 1$. Similarly, if, for some $b > 0$,

$$f_C(\tau) = o(\tau^b), \quad (3.3.5)$$

then a polynomial rate of convergence, faster than $n^{(b-\zeta)/(2b+1)}$ for any $\zeta \in (0, b)$, is possible. Table 3.2 below provides examples of copulas of $(\eta, -\varepsilon)$ satisfying the latter condition (see Appendix B.3.7 for its verification in each case). It is worth noting that for the last two copulas considered in the table, we actually establish that $f_C(\tau)$ tends to zero exponentially fast in $\tau$. In such situations, $(3.3.5)$ holds for all $b$, and it is possible to achieve a rate of convergence for $\hat{\delta}$ and $\hat{\beta}$ that is faster than $n^{1/2-\zeta}$ for any $\zeta > 0$. In other words, an adequate choice of $\tau_n$ can make the rate of convergence arbitrarily close to the standard parametric root-n rate. In all cases, the general idea is that if the tail dependence between $\varepsilon$ and $\eta$ is strong, which can be interpreted as a strong form of endogenous selection, then $f(.)$ is small. It follows that a large $\tau_n$ is admissible, resulting in fast convergence rates.
Table 3.2: Examples of copulas leading to a polynomial rate of convergence.

<table>
<thead>
<tr>
<th>Copula family</th>
<th>Restriction ensuring (3.3.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian $C(u, v; \rho)$</td>
<td>$\rho &gt; 0$</td>
</tr>
<tr>
<td>Clayton $C(u, v; \theta) = \max([u^{-\theta} + v^{-\theta} - 1]^{-1/\theta}, 0)$</td>
<td>$\theta &gt; 0$</td>
</tr>
<tr>
<td>Rotated Gumbel-Barnett $C(u, v; \theta) = u - u(1 - v) \exp(-\theta \log(u) \log(1 - v))$</td>
<td>$\theta \in (0, 1]$</td>
</tr>
<tr>
<td>$C(u, v; \theta) = (1 + [(u^{-1} - 1)^{\theta} + (v^{-1} - 1)^{\theta}]^{1/\theta})^{-1}$</td>
<td>$\theta &gt; 1$</td>
</tr>
<tr>
<td>$C(u, v; \theta) = (1 + [(u^{-1/\theta} - 1)^{\theta} + (v^{-1/\theta} - 1)^{\theta}]^{1/\theta})^{-\theta}$</td>
<td>$\theta \geq 1$</td>
</tr>
<tr>
<td>$C(u, v; \theta) = \theta / \log(\exp(\theta/u) + \exp(\theta/v) - \exp(\theta))$</td>
<td>$\theta &gt; 0$</td>
</tr>
<tr>
<td>$C(u, v; \theta) = [\log(\exp(u^{-\theta}) + \exp(v^{-\theta} - e)]^{-1/\theta}$</td>
<td>$\theta &gt; 0$</td>
</tr>
</tbody>
</table>

Quantile effects

In order to conduct inference on the quantile effects, one needs to (pre)test the homoskedasticity restriction $\delta_j = 0$. Valid inference requires that the critical value of the corresponding t-test depend on the sample size $n$, so that the level of the test tends to zero while the power tends to one. A possibility is to choose the critical values $c_n$ so that $c_n \to \infty$, but slowly enough so that $c_n/\sqrt{n} \to 0$. In practice, we use in our application $c_n = \sqrt{\log(n)}$, which is advocated in different contexts by Andrews (1999) and Andrews and Soares (2010).

We first consider the partially homoskedastic setting. Recall that in this case the quantile effects of $\{X_{jk}\}_{k=1}^{d_3}$, where $\{\delta_{jk}\}_{k=1}^{d_3}$ are equal to zero, are confounded with the average effects of the corresponding covariates. As before, we let $G_{\Delta} = -I_{J+1} \otimes I_{d_3}$, $\Gamma_2 = (0, I_d)$ and $\Gamma_3 = \text{diag}(1, 1/\sqrt{t_1}, \cdots, 1/\sqrt{t_J})$. Finally, we let $\tilde{\lambda}_n = \tilde{\gamma}(\tau_n) \log(m)/\tilde{\gamma}(m\tau_n) - \tilde{\gamma}((\tau_n))$, where $m \neq 1$ denotes an arbitrary positive constant.

Theorem 3.3.2. Under Assumptions 19-21 and 24-27, $\{\delta_{jk}\}_{k=1}^{d_3}$ are zeros and if $W_1 \to^p W_1$, where $W_1$ is symmetric positive definite and nonstochastic, then

$$\tilde{\lambda}_n \sqrt{\tau_n} (\tilde{\Delta}^1 - \Delta^1) \rightsquigarrow N(0, \Delta_{\Omega}),$$

92
where $|\hat{\lambda}_n| \xrightarrow{p} \infty$ and

$$
\Omega_\Delta = (G_\Delta' W_1 G_\Delta)^{-1} G_\Delta' W_1 (\Gamma_3 \otimes \Psi \Gamma_2) (L \otimes \Omega_0) (\Gamma_3' \otimes \Gamma_2' \Psi') W_1 G_\Delta (G_\Delta' W_1 G_\Delta)^{-1}.
$$

The optimal weighting matrices is then $W_\Delta^* = [((\Gamma_3 \otimes \Psi \Gamma_2) (L \otimes \Omega_0) (\Gamma_3' \otimes \Gamma_2' \Psi'))]^{-1}$, and the corresponding asymptotic variances is $(G_\Delta' W_\Delta^* G_\Delta)^{-1}$. Finally, the bootstrap is consistent for $\hat{\Delta}^1$.

Theorem 3.3.2 shows first that $\hat{\Delta}^1$ is asymptotically normal and has a faster rate of convergence than the one of the unconstrained estimator $\hat{\beta}$. It also shows that the bootstrap is consistent for $\hat{\Delta}^1$. In practice, we use bootstrap to conduct inference on the location and scale parameters as well as on the average effects, as it proves to be more reliable than the asymptotic approximation in finite sample.\(^{10}\)

Finally, for the components $j$ such that we reject $\delta_j = 0$, we propose to construct a confidence interval on the quantile effects $\Delta_{j\tau}$ based on the outer bounds given by (3.2.3). Importantly, because the bounds are defined as supremum and infimum of functions that have to be estimated, we can apply the methodology developed by Chernozhukov et al. (2013b) to conduct inference on intersection bounds. We illustrate the procedure by focusing on the upper bound of the confidence interval, when $\delta_j > 0$. The lower bound of the interval and the case $\delta_j < 0$ can be treated similarly.

By Theorem 3.3.1, the rates of convergence of $\hat{\delta}$ and $\hat{\beta}$ are $(\tau_n n)^{-1/2}$ and $\gamma(\tau_n n)^{-1/2}$, respectively. The rates of convergence of $\hat{Q}_{Y|D=1,X=x}$ and $\hat{P}(D = 1|X = x)$ depend on the number of continuous components of $X$, on the degree of smoothness of $x \mapsto Q_{Y|D=1,X=x}$ and $x \mapsto P(D = 1|X = x)$ as well as on the choice of the tuning parameters. In any case, it is always possible to choose $\tau_n$ so that the rate of

\(^{10}\) Results of simulations comparing inference based on bootstrap versus asymptotic variance are available from the authors upon request.
convergence of $\hat{\beta}$ will be slower than the rates of convergence of $\hat{\delta}$, $\tilde{Q}_{Y|D=1,X=x}$ and $\tilde{P}(D = 1|X = x)$.\footnote{This is the case for any $\tau_n$ satisfying Assumption 27 if all the components of $X$ are discrete. If one component of $X$ is continuous, the rate of convergence for $Q_{Y|D=1,X=x}$ and $P(D = 1|X = x)$ will typically be $n^{-2/5}$. Then one has to impose, in addition to Assumption 27, that $\tau_n = o(n^{-1/5})$. Under these conditions, one can show that such a $\tau_n$ always exist by a simple monotonicity argument.} In this case,

$$\theta_j(x) - \hat{\theta}_j(x) = f_j(x)(\beta - \hat{\beta}) + o_P\left(\frac{\gamma(\tau_n)}{\sqrt{\tau_n n}}\right),$$

where $f_j(x) = -e'_j + \frac{\hat{\delta}_j x'}{1 + x' \delta}$ and $e_j$ is a vector of $\mathbb{R}^d$ whose $j$th coordinator equals 1 and others equal 0. One can then apply the inference procedure discussed in Section 4.1 of Chernozhukov et al. (2013b) to construct the upper bound of the confidence interval. Note that although the rate of convergence is not $\sqrt{n}$ here, their Theorem 4.1 still applies, after replacing $\sqrt{n}$ by $\sqrt{\tau_n n}/\gamma(\tau_n)$.\footnote{For that purpose, we need to assume their Condition V, which is a mild regularity condition (see (Chernozhukov et al., 2013b), p.691, for a discussion). Then, using the proof of Theorem 3.3.1, we can check that under Assumptions 19-21 and 24-27, their Conditions P-(ii)-(v) hold (replacing $\sqrt{n}$ by $\sqrt{\tau_n n}/\gamma(\tau_n)$). Although Condition P-(i) does not hold, we can still prove their Lemma 4 using the fact that the nonparametric part of $\theta(\cdot)$ does not play any role in the asymptotic distribution of $\hat{\theta}(\cdot)$.}

Specifically, let us define

$$s_n(x) = \frac{\gamma(\tau_n)}{\sqrt{\tau_n n}} \left\| f_j(x) \hat{\Omega}^{1/2} \right\|_2, \quad Z^*_n(x) = \frac{f_j(x) \hat{\Omega}^{1/2}}{\left\| f_j(x) \hat{\Omega}^{1/2} \right\|_2} N_d.$$  

where $\left\| \cdot \right\|_2$ denotes the Euclidean norm, $\hat{\Omega}_\delta$ is the consistent estimator of the asymptotic variance matrix $\Omega_\delta$ described after Theorem 3.3.1 and $N_d$ is a $d$-dimensional standard normal vector generated independently from the data. Then one can compute, typically by simulations,

$$K_{1n} = Q_{\sup_{x \in \text{Supp}(X)} Z^*_n(x)|\text{data}}(1 - 0.1/\log(n)).$$

Now, constructing $\tilde{X}_n$ as

$$\tilde{X}_n = \left\{ x \in \text{Supp}(X) : \hat{\theta}_j(x) \leq 2K_{1n}s_n(x) + \inf_{\tilde{x} \in \text{Supp}(X)} \tilde{\theta}_j(\tilde{x}) + K_{1n}s_n(\tilde{x}) \right\}.$$
one can compute

\[ K_{2n}(\tau) = Q_{\sup x, \bar{x}_n} z_n^*(x) | \text{data}(\tau). \]

Finally, the upper bound \( \hat{\Delta}_{j,1-\alpha} \) of a confidence interval on \( \Delta_j \) of nominal coverage \( 1 - \alpha \) is defined by

\[ \hat{\Delta}_{j,1-\alpha} = \inf_{x \in \text{Supp}(X)} \hat{\theta}_j(x) + K_{2n}(1 - \alpha) s_n(x). \]

### 3.3.3 Choice of the quantile index

The estimators of \( \beta \) and \( \delta \) are asymptotically normal with zero mean provided that they are based on a sequence of quantile indices \( \tau_n \) satisfying the bias-variance trade-off of Assumption 27. Though there always exists a sequence \( \tau_n \) satisfying Assumption 27 under Assumption 21, admissible rates of convergence towards 0 for \( \tau_n \) are unknown, since they depend on \( f(\gamma(\tau_n)) \), which is itself unknown. A related issue arises in the estimation at infinity of the intercept of sample selection models (see Andrews and Schafgans (1998) and Schafgans and Zinde-Walsh (2002)) or in the estimation of extreme value index (see Drees and Kaufmann (1998) and Danielsson et al. (2001)). We propose in the following a heuristic data-driven method, which consists of selecting \( \tau_n \) as the minimizer of a criterion function that represents the trade-off between bias and variance. The innovative idea here is to combine a subsampling method with a minimum distance estimator to produce a proxy of the bias.

Specifically, let us consider the \( J \) test statistic

\[ T_J(\tau) = \frac{\log(m)^2 \tau n}{(\hat{\gamma}(m\tau) - \hat{\gamma}(\tau))^2} g_n(\hat{\delta}(\tau))' \tilde{W}^*_\delta g_n(\hat{\delta}(\tau)), \]

for some arbitrary fixed \( m > 1 \). Here \( \hat{\delta}(\tau) \) is the estimator obtained using the quantile index \( \tau \) and \( \tilde{W}^*_\delta \) is an estimator of \( W^*_\delta \) using \( \hat{\delta}(\tau) \). We prove in Appendix B.1 that if
\( \tau_n \) satisfies Assumption 27, \( T_J(\tau_n) \) converges to a chi-square distribution with \((J-1)d\) degrees of freedom as \( n \) grows to infinity. We also show that otherwise, the asymptotic distribution of the \( J \) test statistic includes an additional term. Heuristically, this suggests in particular that if the median of the \( J \) test statistic is close enough to the median of a chi-square distribution with \((J-1)d\) degrees of freedom, denoted by \( M_{(J-1)d} \), then the bias term should be small. Our data-driven procedure builds on this idea.

In practice, we propose to estimate the difference between the two medians using subsampling. For each subsample and each quantile index \( \tau \) within a grid defined below, we compute \( T_J(\tau) \). Then, letting \( M_s(\tau) \) denote the median of these test statistics over the different subsamples and for a given \( \tau \), we compute

\[
\hat{\text{diff}}_n(\tau) = \frac{|M_s(\tau) - M_{(J-1)d}|}{\sqrt{b_n \tau}},
\]

where \( b_n \) denotes the subsample size.

Similarly, the asymptotic variance is estimated by the variance of the subsampling point estimates of \( \delta \) multiplied by the normalizing factor \( b_n/n \). We call this estimator \( \hat{\text{Var}}_n(\tau) \). At the end, we select the quantile index as follows:

\[
\hat{\tau}_n = \arg \min_{\tau} \hat{\text{Var}}_n(\tau) + \hat{\text{diff}}_n(\tau).
\]

We thus base our procedure on the trade-off between the variance and our proxy of the bias. It follows that we achieve undersmoothing in comparison with a more standard trade-off between variance and squared bias. Note that, similarly to the case of nonparametric regressions, this is needed to control the asymptotic bias that would otherwise affect the limiting distribution of our estimator.

We implement this method by searching over a grid of \( \tau \) on an interval. We set the upper bound of this interval to 0.3 and the lower bound to \( \min(0.1, 80/b_n) \).
This lower bound is motivated by the fact that if the effective subsampling size $\tau b_n$ becomes too small, then the intermediate order asymptotic theory is likely to be a poor approximation (see Chernozhukov and Fernandez-Val (2011) for a related discussion). Finally, we select the quantile indices for estimating $\Delta^1$ in the partially homoskedastic case in the same manner.

3.4 Simulations

In this section, we investigate the finite-sample performances of our estimation procedure by simulating the following model for four different sample sizes ($n = 250$, $n = 500$, $n = 1,000$ and $n = 2,000$):

$$Y^* = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + (1 + \delta_1 X_1 + \delta_2 X_2 + \delta_3 X_3) \varepsilon$$

$$D = \mathbb{1} \{0.6 + Y^* + 0.3 X_1 + 0.2 X_2 + X_3^2 + \eta \geq 0\}.$$  

$X_1$ and $X_2$ are two mutually exclusive binary variables, such that $X_1 = \mathbb{1}\{U \leq 0.3\}$ and $X_2 = \mathbb{1}\{U \geq 0.8\}$, with $U$ uniformly distributed over $[0, 1]$. $X_3$ is drawn from a truncated normal distribution with support $[-1.8, 1.8]$, mean 0 and standard deviation 1. $(\varepsilon, \eta)$ are jointly normally distributed, with mean zero and covariance matrix $\begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}$. Finally, the true values of the parameters are given by: $\beta_1 = 0.2$, $\beta_2 = 0.4$, $\beta_3 = 0.5$, $\delta_1 = 0$, $\delta_2 = 0.1$ and $\delta_3 = -0.3$.

We report in Table 3.3 below, for each sample size, the bias and standard deviation for eight different estimators. Namely, we first estimate $(\delta_1, \delta_2, \delta_3)$ and $(\beta_1, \beta_2, \beta_3)$ without imposing $\delta_1 = 0$. Then we impose $\delta_1 = 0$ and estimate $\beta_1$ under this homoskedasticity constraint. As shown in Section 3.2.2 (Theorem 3.3.2), $\beta_1$ is estimated at a faster convergence rate in the latter case. We use the two-step, asymptotically optimal estimators of $\beta$ and $\delta$. We also document the severity of the selection bias in this context by reporting the bias and standard deviation of a naive OLS estimator.
of $\beta_1$ only using the observations such that $D = 1$. Throughout this section we pay special attention to the performances of our estimator of $\Delta^1$ since, in our application, the black-white wage gap will be estimated similarly.

The vector of spacing parameters $l_j$ used in minimum distance estimation is set equal to $(0.65, 0.85, 1.15, 1.45)$. Intuitively, these parameters have to differ sufficiently to provide enough variation. At the same time, they should not be too large nor too small, otherwise the corresponding quantile indices $\tau_n l_j$ might escape from the intermediate order region. However, in practice, our estimates do not appear to be meaningfully sensitive to the choice of $l_j$. The choice of the quantile index $\tau_n$ is more critical. We choose this parameter using the data-driven method discussed in Section 3.3.3, with subsample sizes $(150, 300, 500, 700)$ corresponding to the four total sample sizes $(250, 500, 1,000$ and $2,000$) and 500 subsamples in each case. We report in Table 3.3 below the average quantile indices computed across all simulations.
<table>
<thead>
<tr>
<th>True Value</th>
<th>( \delta_1 )</th>
<th>( \delta_2 )</th>
<th>( \delta_3 )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \beta_3 )</th>
<th>( \Delta \gamma (= \beta_1) )</th>
<th>OLS</th>
<th>( \beta_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>n=250</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>0.077</td>
<td>0.077</td>
<td>0.109</td>
<td>-0.053</td>
<td>-0.049</td>
<td>-0.071</td>
<td>0.001</td>
<td>0.090</td>
<td></td>
</tr>
<tr>
<td>Std dev</td>
<td>0.342</td>
<td>0.419</td>
<td>0.152</td>
<td>0.255</td>
<td>0.308</td>
<td>0.104</td>
<td>0.196</td>
<td>0.096</td>
<td></td>
</tr>
<tr>
<td>Average ( \tau_n )</td>
<td>0.261</td>
<td></td>
<td></td>
<td>0.261</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.205</td>
</tr>
<tr>
<td>n=500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>0.063</td>
<td>0.092</td>
<td>0.064</td>
<td>-0.032</td>
<td>-0.071</td>
<td>-0.040</td>
<td>0.012</td>
<td>0.085</td>
<td></td>
</tr>
<tr>
<td>Std dev</td>
<td>0.269</td>
<td>0.341</td>
<td>0.118</td>
<td>0.208</td>
<td>0.276</td>
<td>0.090</td>
<td>0.135</td>
<td>0.066</td>
<td></td>
</tr>
<tr>
<td>Average ( \tau_n )</td>
<td>0.225</td>
<td></td>
<td></td>
<td>0.225</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.204</td>
</tr>
<tr>
<td>n=1,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>0.042</td>
<td>0.028</td>
<td>0.043</td>
<td>-0.023</td>
<td>-0.015</td>
<td>-0.029</td>
<td>0.001</td>
<td>0.083</td>
<td></td>
</tr>
<tr>
<td>Std dev</td>
<td>0.198</td>
<td>0.235</td>
<td>0.087</td>
<td>0.178</td>
<td>0.209</td>
<td>0.071</td>
<td>0.095</td>
<td>0.042</td>
<td></td>
</tr>
<tr>
<td>Average ( \tau_n )</td>
<td>0.206</td>
<td></td>
<td></td>
<td>0.206</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.204</td>
</tr>
<tr>
<td>n=2,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>0.006</td>
<td>0.019</td>
<td>0.021</td>
<td>-0.006</td>
<td>-0.014</td>
<td>-0.016</td>
<td>-0.009</td>
<td>0.082</td>
<td></td>
</tr>
<tr>
<td>Std dev</td>
<td>0.125</td>
<td>0.164</td>
<td>0.061</td>
<td>0.115</td>
<td>0.155</td>
<td>0.056</td>
<td>0.068</td>
<td>0.033</td>
<td></td>
</tr>
<tr>
<td>Average ( \tau_n )</td>
<td>0.188</td>
<td></td>
<td></td>
<td>0.188</td>
<td></td>
<td></td>
<td></td>
<td>0.206</td>
<td></td>
</tr>
</tbody>
</table>

Note: Results were obtained using 280 simulations for each sample size.

Importantly, for each sample size, the bias-standard deviation ratio for each estimator is much smaller than 1, consistent with our data-driven choice of \( \tau_n \) leading to undersmoothing. Besides, the standard deviations of our estimators as well as the average \( \tau_n \) generally decrease with the sample size, as expected given the consistency of our estimators and the bias-variance tradeoff underlying the choice of \( \tau_n \). In practice, our estimators exhibit a fairly small bias for sample sizes larger than \( n = 500 \). Note also that the constrained version of the estimator of \( \beta_1 \), which makes use of the homoskedasticity constraint \( \delta_1 = 0 \), is much more precise than the unconstrained estimator. The OLS estimator of \( \beta_1 \), on the other hand, displays a large bias for all sample sizes.
The average \( \tau_n \) that we choose in these simulations are relatively large. This reflects the fact that in this setup, \( D \) strongly depends on \( Y^* \) and, consistent with our application, \( D = 1 \) for a large fraction of observations.\(^{13}\) As discussed in Section 3.2.1, \( f(.) \) is small in this case and the bias term becomes quickly dominated by the variance term. It follows that our data-driven procedure yields relatively large values of \( \tau_n \). However, it is worth noting that our estimators are based on quantile regressions for the quantile indices \( \tau_n l_j \), with \( l_j \in \{0.65, 0.85, 1.15, 1.45\} \). For instance, the smallest quantile indice used in the estimation procedure is equal to 0.13 for \( \tau_n = 0.2 \) (the average \( \tau_n \) for our constrained estimator of \( \beta_1 \)).

Figure 3.1 above displays the Mean Squared Error (MSE) of the estimator \( \hat{\Delta}^1 \)

\(^{13}\) We also ran additional simulations using data generating processes where the dependence between \( D \) and \( Y^* \) is weaker and \( D = 1 \) for fewer observations. Simulation results indicate that the average \( \tau_n \) gets close to the lower bound of the grid (0.1) in those cases
as a function of the quantile index \( \tau_n \). The plots corresponding to \( n = 1,000 \) and \( n = 2,000 \) exhibit a U-shaped relationship between the MSE and the quantile index. This pattern reflects a bias-variance tradeoff with respect to the choice of \( \tau_n \). When the quantile index is small, the bias is small but the variance is large, and vice versa. On the other hand, the relationship between MSE and \( \tau_n \) is mostly decreasing for \( n = 250 \) and \( n = 500 \). This is consistent with the variance term dominating the bias term for \( \tau_n < 0.3 \) and such small sample sizes. The vertical line corresponds to the average \( \tau_n \) (across simulations) chosen based on our data-driven method. For all sample sizes, this index is smaller than the one yielding the smallest MSE, consistent with our data-driven method tending to undersmooth. However, for sample sizes larger than 250, the MSE evaluated at the average selected quantile index is pretty close to the minimum.
Finally, we examine in Figure 3.2 above the relationship between the coverage of the 95% and 97.5% confidence intervals constructed with our estimator \( \hat{\Delta}^1 \) and the quantile index \( \tau_n \). The confidence intervals are constructed using percentile bootstrap with 500 replications. The coverage is generally quite close to the nominal rates for values of \( \tau_n \) in the neighborhood of the average quantile index obtained with our data-driven method, with the exception of \( n = 500 \) where the confidence intervals tend to overcover. The sharp decline in coverage for larger values of the quantile index for \( n = 1,000 \) and \( n = 2,000 \) reflects the existence of a nonvanishing bias for fixed values of \( \tau_n \). This stresses the importance of carefully choosing the quantile index in order to conduct valid inference on the average effects. These simulation results
indicate that our data-driven procedure does a good job in selecting appropriate quantile indices.

3.5 Application to the black-white wage gap

We apply our method to the estimation of the black-white wage gap among young males for two groups of cohorts, using data from the National Longitudinal Survey of Youth 1979 (NLSY79) and National Longitudinal Survey of Youth 1997 (NLSY97). Individuals surveyed in the NLSY79 were 14 to 22 years old in 1979, while individuals from the NLSY97 were 12 to 16 years old in 1997. In the following, we are interested in estimating the black-white wage gap for these two groups of individuals as of 1990-1991 and 2007-2008, respectively. As noted in early articles by Butler and Heckman (1977) and Brown (1984), and documented more recently by Juhn (2003), among males, blacks are significantly more likely to dropout from the labor market. To the extent that those dropouts tend to have lower potential wages, it follows that failure to control for endogenous labor market participation is likely to result in underestimating the black-white wage differential. It is worth noting that finding a valid instrument for selection is particularly difficult in the context of male labor force participation. As a result, most of the attempts to deal with selection have consisted of imputing wages for non-workers (see, among others, (Brown, 1984), (Smith and Welch, 1989), (Neal and Johnson, 1996), (Juhn, 2003), (Neal, 2004), (Neal, 2006) and (Neal and Rick, 2014)).

Importantly, since across-cohort changes in selection into the workforce is also different for blacks and for whites, adequately dealing with selection is needed to obtain credible estimates of the across-cohort evolution of the black-white wage gap. Altonji and Blank (1999) stress the importance of correcting for changes in race differential selection into work, and review some of the empirical literature addressing
3.5.1 Evidence from the NLSY79

We first use our method to estimate the black-white wage gap among young males from the NLSY79, revisiting the influential work of Neal and Johnson (1996) on this question. We use the same sample as Neal and Johnson (1996) in our analysis, and consider as they did that an individual is a nonparticipant if he did not work in 1990 nor in 1991. The total sample size is $n = 1,674$, with an overall labor force participation rate over the period of interest (1990-1991) equal to 95%. We refer the reader to Neal and Johnson (1996) for a detailed discussion on the data.

We start by replicating the results of Neal and Johnson (1996) in Table 3.4 below by running four regressions on the log of hourly wages on a set of observable characteristics, namely black, Hispanic dummies and age (specifications (1) and (3)), together with AFQT and AFQT squared (specifications (2) and (4)). The first two columns contain the results of simple OLS regressions, replicating Columns (1) and (3) in Table 1 of Neal and Johnson (1996) (p.875), while in the last two columns we replicate their Table 4 (p.883) by imputing a zero log-wage for nonparticipants and running a median log-wage regression. As discussed in Neal and Johnson (1996) and more extensively in Johnson et al. (2000), this imputation method yields consistent estimates under the assumption that, conditional on the set of observable characteristics included in the regression, the potential wage for any individual who did not work neither in 1990 nor in 1991 lies below the median. It is important to note that the identifying condition of independence at infinity used in our paper (Assumption

---

14 As the authors put it, “Comparisons of average or median wages of persons with jobs do not provide an accurate picture of changes in the offer distributions faced by black and by white workers” (pp. 3240). See also Juhn (2003), who provides evidence that the evolution over the period 1969-1998 of the black-white wage gap is severely biased if one does not take into account the decline in work participation rates of black men relative to white men. In recent work, Neal and Rick (2014) show that the growth in prison populations in the last decades is an important factor behind the evolution of differential workforce participation of blacks and whites.
21) relaxes this assumption by replacing the median with some extremal quantile of the conditional wage distribution.\textsuperscript{15} As is put forward by Neal and Johnson (1996), Columns (1) and (2) show that the estimated black-white wage gap drops sharply, from 24.4% to 7.1%, after adding controls for ability, namely AFQT and AFQT squared. It is also worth noting that the estimated black-white wage differential changes substantially, increasing (in absolute value) by as much as 6.4 points, after addressing the selection issue with the imputation method proposed in Neal and Johnson (1996) (see Columns (2) and (4)).

\begin{table}[h]
\centering
\begin{tabular}{|l|c|c|c|c|}
\hline
 & (1) & (2) & (3) & (4) \\
\hline
Black & -0.244 & -0.071 & -0.356 & -0.135 \\
 & (0.026) & (0.027) & (0.028) & (0.034) \\
Hispanic & -0.114 & 0.005 & -0.181 & -0.013 \\
 & (0.030) & (0.030) & (0.033) & (0.038) \\
Age & 0.048 & 0.040 & 0.068 & 0.055 \\
 & (0.014) & (0.013) & (0.016) & (0.017) \\
AFQT & — & 0.173 & — & 0.206 \\
 & (0.012) & (0.016) & (0.015) \\
AFQT\textsuperscript{2} & — & -0.013 & — & 0.010 \\
 & (0.011) & (0.011) & (0.014) \\
\hline
\end{tabular}
\caption{OLS and median log-wage regression results (NLSY79)}
\end{table}

Note: Standard errors are reported in parentheses.

We now investigate how the above results are changed when we use our estimation method and implement the two-step asymptotically optimal estimators of $\delta$ and $\beta$. Table 3.5 presents the estimation results for the heteroskedasticity parameters $\delta$ and the parameters $\beta$. Since we fail to reject homoskedasticity for all the covariates with the exception of age, we report both the corresponding unconstrained (“Heteroskedastic”) and constrained (“Homoskedastic”) estimates of $\beta$. In the discussion below we focus on our preferred constrained estimates, which have a structural interpretation in terms of average effects.

\textsuperscript{15} Our identifying condition is also weaker in the sense that $h$ does not need to be equal to 1.
Table 3.5: Extremal quantile regression results (NLSY79)

<table>
<thead>
<tr>
<th></th>
<th>Heteroskedastic</th>
<th>Homoskedastic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\delta}$</td>
<td>$\hat{\beta}$</td>
</tr>
<tr>
<td>Black</td>
<td>0.019</td>
<td>-0.215</td>
</tr>
<tr>
<td></td>
<td>(0.059)</td>
<td>(0.335)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.005</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>(0.059)</td>
<td>(0.335)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.029</td>
<td>0.215</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.032)</td>
</tr>
<tr>
<td>AFQT</td>
<td>-0.005</td>
<td>0.238</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.116)</td>
</tr>
<tr>
<td>AFQT$^2$</td>
<td>-0.007</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.127)</td>
</tr>
</tbody>
</table>

Note: Bootstrap standard errors (500 replications) are reported in parentheses. We perform the homoskedasticity tests using the critical values $c_n = \sqrt{\log(n)}$, where $n$ is the sample size ($n = 1,674$ here). The vector of spacing parameters $l_j$ used in minimum distance estimation is equal to $(0.65, 0.85, 1.15, 1.45)$. The quantile index $\tau_n$ is chosen based on the data-driven procedure discussed in Section 3.3, using 500 subsamples of size 550.

The estimation results from our extremal quantile method show that the size of the black-white wage gap (10.1%) is smaller than the estimated gap obtained under the imputation method proposed by Neal and Johnson (1996) (13.5%), but larger than the gap estimated using simple OLS (7.1%). The fact that our preferred estimate of the black-white wage gap is smaller than the one obtained with the imputation method is consistent with our estimator being based on a weaker identifying assumption. While Neal and Johnson (1996) assume that, conditional on observed characteristics, those individuals who do not participate to the labor market have a potential wage below the median, a sufficient condition to apply our method is to rule out the possibility that non-participants have arbitrarily large potential wages. Intuitively it follows that our approach results in a milder form of selection correction, which is consistent with our findings.

Finally, it is worth stressing that our results are in line with the key takeaway
of Neal and Johnson (1996), namely that premarket factors, as measured here by AFQT, account for most of the black-white wage differential. In fact, our results point to an even more important role played by AFQT, since the estimated wage gap drops from close to the median regression estimate (around 35%) to 10.1% after adding AFQT.\footnote{Estimation results from our method without controlling for AFQT are not reported here to save space. They are available from the authors upon request.}

### 3.5.2 Across-cohort evolution

We now examine the evolution across the NLSY79 and NLSY97 cohorts of the black-white wage gap. To do so, we apply our method to estimate the wage gap using hourly wages measured in 1990-1991 for the NLSY79 sample and in 2007-2008 for the NLSY97 sample. We follow Altonji et al. (2012) by using a modified version of the AFQT variable, which corrects for the across-cohort changes in the ASVAB test format as well as in the age ranges at which the test was taken. This age correction procedure is based on an equipercentile mapping. To the extent that the rank within the AFQT distribution may vary with the age of the respondent at the time of the test, we further restrict the samples to the respondents who took the test when they were 16 or 17. Besides this age restriction, we constructed the NLSY97 sample so as to match as closely as possible the sample selection rules used by Neal and Johnson (1996) for the NLSY79. Consistent with prior evidence, we find that the labor force participation rate of black men has fallen over time relative to white men (see Appendix B.2 for more details on the data). The baseline estimation results are reported in Table 3.6 below. The resulting sample sizes are equal to 1,077 and 1,123 for the NLSY79 and NLSY97 cohorts, respectively.
Table 3.6: Extremal quantile and median regression results (NLSY79-NLSY97)

<table>
<thead>
<tr>
<th></th>
<th>NLSY79</th>
<th></th>
<th>NLSY97</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Extremal Quantile</td>
<td>Median</td>
<td>Extremal Quantile</td>
<td>Median</td>
</tr>
<tr>
<td>Black</td>
<td>-0.122 (0.044)</td>
<td>-0.145 (0.039)</td>
<td>-0.140 (0.057)</td>
<td>-0.167 (0.058)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.029 (0.055)</td>
<td>-0.017 (0.056)</td>
<td>-0.054 (0.050)</td>
<td>-0.089 (0.050)</td>
</tr>
<tr>
<td>AFQT</td>
<td>0.185 (0.020)</td>
<td>0.180 (0.019)</td>
<td>0.153 (0.025)</td>
<td>0.111 (0.026)</td>
</tr>
<tr>
<td>AFQT^2</td>
<td>0.007 (0.017)</td>
<td>0.008 (0.017)</td>
<td>0.002 (0.020)</td>
<td>-0.023 (0.020)</td>
</tr>
</tbody>
</table>

Notes: Estimations also include linear control for age. Bootstrap standard errors (500 replications) are reported in parentheses. In the column “Extremal Quantile”, we report the results corresponding to our preferred constrained specification, since we fail to reject homoskedasticity for all of the covariates with the exception of age. We perform the homoskedasticity tests using the critical values $c_n = \sqrt{\log(n)}$, where $n$ is the sample size. The vector of spacing parameters $l_j$ used in minimum distance estimation is equal to $(0.65, 0.85, 1.15, 1.45)$. The quantile index $\tau_n$ is chosen based on the data-driven procedure discussed in Section 3.3, using 500 subsamples of size 500.

The estimation results obtained with our method (“Extremal Quantile” columns) provide evidence of a wider black-white wage gap for the 1997 cohort relative to the 1979 cohort, with an increase in the estimated gap from 12.2% to 14%. It is also interesting to note that, while the estimated levels do differ across both methods, the results from the median regression of Neal and Johnson (1996) (“Median” columns) imply an across-cohort increase of a similar magnitude (from 14.5% to 16.7%).

It is important to step back and try and understand what these results really mean. Specifically, do they suggest that labor market discrimination against blacks has actually gotten worse over the last two decades? Or does the estimated increase in the black-white wage gap reflect the fact that the AFQT score only captures a fraction of all the premarket factors that matter on the labor market, which may have changed over time? In particular, the results reported in Table 3.6 provide clear evidence of a decline across cohorts in the wage returns to AFQT, consistent with the latter story. Recent work by Castex and Dechter (2014) also provides evi-
idence from the NLSY79 and NLSY97 that the wage returns to AFQT have decreased over time (see also (Beaudry et al., 2013), who argue that there has been a decline in the demand for cognitive skills in the U.S. since 2000). While providing a definite answer to those questions is particularly challenging, we attempt to shed light on this issue by controlling for additional premarket factors, namely parental education and household structure (as measured by the presence of both biological parents at age 14). Bringing those characteristics into the analysis is important since differences in family environment have been found to account for most of the black-white gap in noncognitive skills (see, e.g., (Carneiro et al., 2005)).

Table 3.7 below reports the estimated black-white wage gap for the 1979 and 1997 cohorts, using our extremal quantile method and the median regression of Neal and Johnson, for three different specifications. The first specification ("No premarket factors") only controls for age and the Hispanic dummy, the second specification ("AFQT only") also controls for AFQT and AFQT squared, while the third specification ("Preferred") further controls for parental education and household structure.

<table>
<thead>
<tr>
<th></th>
<th>NLSY79 Extremal Quan</th>
<th>Median</th>
<th>NLSY97 Extremal Quan</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black (No prem.)</td>
<td>-0.342 (0.041)</td>
<td>-0.349 (0.032)</td>
<td>-0.296 (0.055)</td>
<td>-0.311 (0.051)</td>
</tr>
<tr>
<td>Black (AFQT)</td>
<td>-0.122 (0.044)</td>
<td>-0.145 (0.039)</td>
<td>-0.140 (0.057)</td>
<td>-0.167 (0.058)</td>
</tr>
<tr>
<td>Black (Preferred)</td>
<td>-0.099 (0.045)</td>
<td>-0.123 (0.042)</td>
<td>-0.087 (0.075)</td>
<td>-0.135 (0.064)</td>
</tr>
</tbody>
</table>

Notes: Bootstrap standard errors (500 replications) are reported in parentheses. The “preferred” specification includes AFQT, parental education and household structure. For that case, the sample is then restricted to the individuals with non-missing parental education and household structure, resulting in sample sizes equal to 1,016 for the NLSY79 and 1,071 for the NLSY97. In the column “Extremal Quan”, we report the results corresponding to our preferred constrained specification, since we fail to reject homoskedasticity for the black dummy. We perform the homoskedasticity tests using the critical values $c_n = \sqrt{\log(n)}$, where $n$ is the sample size. The vector of spacing parameters $l_j$ used in minimum distance estimation is equal to (0.65, 0.85, 1.15, 1.45). The quantile index $\tau_n$ is chosen based on the data-driven procedure discussed in Section 3.3, using 500 subsamples of size 500.
Without controlling for premarket factors, our estimation results show that the black-white wage gap has decreased by 4.6 points across the 1979 and 1997 cohorts. This result provides evidence of a very slow black-white wage convergence between 1990 and 2007. While most of the available evidence in the literature relates to the evolution of the black-white wage gap before 2000, it is interesting to note that our results are of the same order of magnitude as the estimates obtained by Neal and Rick (2014) using different datasets (namely the Census Long Form for the year 1990 and the American Community Survey for the year 2007). In their paper, Neal and Rick address the issue of differential selection into the workforce by examining the sensitivity of the median black-white wage gap to various imputation rules, which vary based on the fraction of (missing) wages supposed to fall below the median of the potential wage distribution. This type of sensitivity analysis cannot be used after adding controls for premarket factors, since in that case knowing the fraction of wages falling below or above the median is not sufficient to estimate the median wage gap.

While we find that the black-white wage gap increases over time after controlling for AFQT, Table 3.7 shows that the direction of the change is overturned when including other premarket factors in addition to the AFQT. Using our estimation method, the black-white wage gap is found to be fairly stable across cohorts, declining by only 1.2 points (from 9.9% to 8.7%) between 1990 and 2007. This result suggests that the across-cohort increase in the wage gap conditional on AFQT is actually attributable to the premarket factors other than AFQT, thus reflecting a time-varying omitted variable bias based on these family environment characteristics. Interestingly, one can understand this result as extending the key finding of Neal and Johnson (1996) to the across-cohort change in the wage gap. Premarket factors are a dominant component of the black-white wage gap, not only in level but
also in evolution.

In sum, our estimation results provide evidence of (i) a slow convergence in the raw male black-white wage gap between 1990 and 2007 (Specification without premarket factors), and (ii) an even slower convergence in the residual portion of the black-white wage gap, which remains after controlling for premarket factors (Preferred specification). While we do find that differences in premarket factors are a key component of the black-white wage gap and, as such, should be a major focus from a policy standpoint, the fact that its residual portion remains virtually stable after almost 20 years is also concerning.

We conclude this section by examining whether one could alternatively estimate the across-cohort evolution of the black-white wage gap by applying the inverse density weighting scheme of Lewbel (2007), treating AFQT as a special regressor. Note that, in this context, AFQT appears to be the only potential candidate as a special regressor, thus ruling out the possible use of the special regressor method in the absence of controls for premarket skills. The large support condition would require the employment probability to be arbitrarily small for some values of the AFQT. Although there is some variation, we found that the conditional employment probability, estimated via nonparametric regression, remains very far from 0, specifically above 0.63 for both NLSY cohorts. This clearly indicates that this method could not be used in this context.

3.6 Concluding remarks

In this paper, we develop a new semiparametric inference method for location-scale models in the presence of sample selection. A key feature of our method is that it can be used in situations where one does not have access to an instrument for selection, nor to a large support regressor. Instead, the main identifying condition is based
on selection being independent of the covariates for large values of the outcome. We show that this condition is typically mild provided that selection is endogenous. Building on this identification strategy, we propose a simple estimation procedure, which combines quantile regressions in the tails, or extremal quantile regressions, with minimum distance. We establish the consistency and asymptotic normality of our estimators by extending the analysis of Chernozhukov (2005b) to a setting with sample selection, and show that bootstrap is consistent. The choice of an appropriate quantile index is important in this context, and we derive a data-driven procedure for this purpose. Importantly for the practical usefulness of our method, we show that our estimation procedure performs well even with fairly small samples.

Finally, we apply our method to the estimation of the black-white wage gap among males from the NLSY79 and NLSY97 cohorts. Correcting for selection into the workforce is key in this context since black males are more likely to dropout from the labor market than white males, and this difference has increased over time. Our estimation results show that premarket factors play a major role in explaining the magnitude of the black-white wage gap, as well as its evolution over time.
4

\( \sqrt{n} \)-Consistency of the Intercept of a Binary Response Model Under Tail Restrictions

4.1 Introduction

In the seminal works of Lewbel (1997) and Lewbel (2000), the intercept of the binary response model is identified and estimated with the aid of a special regressor \( V \). When \( V \) has compact support, Lewbel (1997), Lewbel (2000), and Lewbel and Schennach (2007) have shown that the estimator is \( \sqrt{n} \)-consistent. But for the sake of identification, the compactness of the support of \( V \) implies that either \( \varepsilon \) also has a compact support (Lewbel, 1997, 2000) or \( \varepsilon \) has tail symmetry (Magnac and Maurin, 2007). The former condition excludes the basic logit and probit models, as pointed out by Lewbel (1997) and Lewbel (2000). The latter condition depends on the unknown intercept value being identified and thus is not generic. When \( V \) has unbounded support, the \( \sqrt{n} \)-consistency has been established based on high-level assumptions on the bias and variance of the estimator. See, for example, Lewbel (1997), Lewbel (2000), and Stoker (1991). These high level assumptions do not hold in general because Khan and Tamer (2010) showed that the intercept is irregularly
identified and cannot be $\sqrt{n}$-consistently estimated without additional tail restrictions. In addition, Khan and Tamer (2010) pointed out that the relative thickness of the tails of $V$ and $\varepsilon$ plays the key role of determining the convergence rate, but they did not provide sufficient conditions for the $\sqrt{n}$-consistency.

This paper provides additional tail restrictions that are sufficient for the $\sqrt{n}$-consistency of the semiparametric estimator of the intercept when both $V$ and $\varepsilon$ are supported on the real line. Given the additional restrictions, we propose a $\sqrt{n}$-consistent estimator as an inverse density-weighted average with trimming based on the extremal quantile of the special regressor $V$. In one particular case, we show that the intercept is $\sqrt{n}$-consistently estimable if the unobservable $\varepsilon$ has rapidly varying tails (e.g. normal distribution) and the special regressor $V$ has regularly varying tails (e.g. T distributions with any degree of freedom). This result extends the previous results in Khan and Tamer (2010) that if $V$ has infinite variance, there exists a $\sqrt{n}$-consistent estimator of the intercept. We also provide sufficient conditions for the non-existence of any $\sqrt{n}$-consistent estimator.

The trimming plays the key role of adjusting the bias-variance tradeoff: if we allow the special regressor to go deep into its tails, we will have a small bias but the density will be close to 0, which makes the inverse density weighting estimator diverge. On the other hand, if we let the special regressor go to the tails slowly, then the bias will also vanish slowly. We show that under additional tail restrictions, trimming based on the extremal quantile of the special regressor can manage to deal with both bias and variance so that the semiparametric estimator can achieve $\sqrt{n}$-consistency.

Asymptotic trimming has already been considered in binary response models by Yang (2015). There are two major differences between our paper and Yang (2015): (1) we focus on $\sqrt{n}$-consistency, whereas he focuses on the applicability of the central
limit theorem; (2) our trimming is based on the extremal quantile of \( V \), whereas his trimming is not driven by data.

Trimming by extremal quantile is not new, either. It is applied to GMM estimators with heavy-tailed data by Hill and Renault (2010) and to average treatment effect estimators by Chaudhuri and Hill (2013). Our method has two main difference: (1) we focus on \( \sqrt{n} \)-consistency, whereas the two aforementioned papers are about the general adaptive convergence rate; (2) both Hill and Renault (2010) and Chaudhuri and Hill (2013) propose to trim the whole estimand, whereas we directly trim the special regressor. The second point is nontrivial because the estimand in our case contains the density of \( V \), which is unknown and should be estimated. Therefore, it is infeasible to directly trim the estimand in our case. What we propose is to trim the special regressor \( V \) and replace the unknown density by its nonparametric estimator.

The rest of the paper is organized as follows. Section 4.2 defines a simple model and proposes to estimate the intercept as an inverse density-weighted average. Section 4.3 shows that the estimator proposed in Section 4.2 is \( \sqrt{n} \)-consistent under additional tail restrictions and asymptotic trimming by extremal quantile of the special regressor. Section 4.4 extends the model to the one considered in Dong and Lewbel (2015), which incorporates endogenous regressors \( X \) and instrumental variables \( Z \). Section 4.5 shows a brief simulation and Section 4.6 concludes. All proofs are collected in the Appendix.

4.2 Definition of the Semiparametric Estimator

The standard binary response model is

\[
Y_i = 1\{X_i'\beta - \varepsilon_i \geq 0\}.
\]
Throughout the paper, we do not impose distributional assumptions on $\varepsilon$. So the model is semiparametric. Identification and estimation of semiparametric binary choice models have been widely investigated in the literature. If covariates $X$ are independent of $\varepsilon$, the average derivative estimators of Powell et al. (1989) and Stoker (1991) can be applied, but only the coefficients for continuously distributed regressors can be estimated. The estimation methods proposed by Han (1987), Ichimura (1993) and Klein and Spady (1993) can handle coefficients for discrete covariates but not the intercept. Furthermore, heteroskedasticity, which is important for economic applications, is ruled out by the full independence assumption. Manski (1975) and Manski (1985) considered the conditional quantile independence, which allows for heteroskedasticity. But the maximum score estimator proposed by the above two papers is neither $\sqrt{n}$-consistent nor asymptotically normal. In addition, Manski (1988) pointed out that a conditional mean restriction does not identify parameter $\beta$. Lewbel (1997) and Lewbel (2000) complement the conditional mean restriction by assuming the existence of a "special regressor" that is independent of $\varepsilon$ given other covariates $X$. Then under certain support conditions, Lewbel (1997) and Lewbel (2000) established the identification of $\beta$ based on conditional mean restriction, allowing for heteroskedasticity on all covariates $X$ except the special regressor. For identification, we will follow the strategy of Lewbel (1997) and Lewbel (2000).

To simplify the discussion, in this section, we follow Khan and Tamer (2010) and consider the case in which $X$ contains only a constant term and a single regressor $V$. The coefficient in front of $V$ is normalized to 1. Then, the simple model can be written as

$$Y_i = 1\{\alpha + V_i - \varepsilon_i \geq 0\}. \quad (4.2.1)$$

Following Lewbel (1997), we make the next assumption.

**Assumption 28.**
(1) \( \{(\varepsilon_i, V_i)\}_{i=1}^{n} \) is i.i.d. (2) \( \varepsilon_i \) and \( V_i \) have full support \( \mathbb{R} \). (3) \( V_i \perp \varepsilon_i \). (4) \( E \varepsilon_i = 0 \).

\( V \) is referred to as the special regressor by Lewbel (1997) and Lewbel (2000) because it is independent of the unobservable \( \varepsilon \) and its support is the real line. Assumption 28(4) is the common location normalization. For the general model considered in Section 4.4, \( X \) includes both the special regressor \( V \) and other covariates that may be endogenous. Assumption 28(3), then, will be relaxed to hold conditionally and Assumption 28(4) will be generalized to an orthogonality condition that allows for heteroskedasticity.

Under Assumption 28, Lewbel (1997) and Lewbel (2000) showed that \( \alpha \) is identified as

\[
\alpha = E \left[ \frac{Y_i - 1\{V_i > 0\}}{f(V_i)} \right],
\]

in which \( f(\cdot) \) denotes the true density of \( V \).

Since the support of \( V \) is the real line, its density vanishes at the tails. Consequently, we face the “zero-denominator” problem. To deal with this problem, we propose an estimator that is the sample analogue of the RHS of Equation (4.2.2) with a trimming function \( \hat{I}_{n,i} \):

\[
\hat{\alpha}_n := \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{Y_i - 1\{V_i > 0\}}{f(V_i)} \right] \hat{I}_{n,i},
\]

in which \( \hat{f} \) is a kernel estimator of \( f \), \( \hat{I}_{n,i} = 1\{V_i \in \hat{S}_n\} \), and \( \hat{S}_n = (\hat{l}_n, \hat{r}_n) \). \( \hat{I}_{n,i} \) is the feasible trimming function whose infeasible counterpart is \( I_{n,i} := 1\{V_i \in S_n\} \), in which \( S_n := (l_n, r_n) \) where \( l_n \) and \( r_n \) are the non-random asymptotic trimming points. \( \hat{l}_n \) and \( \hat{r}_n \) are the estimator for \( l_n \) and \( r_n \), respectively. \( l_n, r_n, \hat{l}_n \) and \( \hat{r}_n \) will be defined in the next section.
4.3 Asymptotic Properties

In this section, we first introduce a set of tail restrictions and clarify the trimming scheme. Under these restrictions, we show that \( \hat{\alpha}_n \) as defined in (4.2.3) is \( \sqrt{n} \)-consistent, asymptotically normal, and semiparametrically efficient. We then introduce another set of tail restrictions that do not intersect with the previous restriction for \( \sqrt{n} \)-consistency. Under the new set of restrictions, we show that there is no regular estimator in existence for \( \alpha \).

**Assumption 29.** Recall that \( f(\cdot) \) is the density of \( V \).

1. \( f(\cdot) \) is monotonic in its left and right tails and is bounded away from 0 on any compact subset of \( \mathbb{R} \).

2. \( f \) is \( \nu \)-th order continuously differentiable with \( \nu \geq 2 \). All of its derivatives are bounded.

3. For some positive constant \( \sigma \), there exists positive constants \( c_1, c_2, \) and \( c_3 \) such that

\[
f(v \pm c_2) \leq c_1 f(v)^{1-\sigma}
\]

when \( |v| \geq c_3 \).

All conditions are mild. Assumption 29(1) and 29(2) are widely used in kernel density estimations. Assumption 29(3) holds when \( f(v) \) decays polynomially as \( |v| \to \infty \) and the density of \( V \) is bounded. To see this, note that by Definition A in Section 0.4.1 of Resnick (2007), if \( f(v) \) decays polynomially\(^1\) as \( |v| \to \infty \), then

\[
\frac{f(v \pm c_2)}{f(v)} \to 1 \quad \text{and} \quad f(v \pm c_2) \leq \sup_v f(v)^\sigma f(v \pm c_2)^{1-\sigma} \leq cf(v \pm c_2)^{1-\sigma} \sim c_1 f(v)^{1-\sigma}.
\]

\(^1\ f(\cdot) \) is regularly varying at \( \infty \).
When \( f(\cdot) \) decays exponentially, as in case of normal density, we further consider the case that \( \log(f(\cdot)) \) decays polynomially. In this case, \( \frac{\log(f(v+c_2))}{\log(f(v))} \to 1 \) as \( |v| \to \infty \).

Thus for any \( \sigma \in (0,1) \) and \( v \) large enough, \( 1 - \sigma \leq \frac{\log(f(v+c_2))}{\log(f(v))} \leq \frac{1}{1-\sigma} \). In addition, when \( f(v) < 1, \log(f(v)) < 0 \). Therefore, we have

\[
\log(f(v \pm c_2)) \leq (1 - \sigma) \log(f(v)), \quad \text{or equivalently,} \quad f(v \pm c_2) \leq f^{1-\sigma}(v).
\]

Next, we state the requirement for the kernel function used to estimate \( f \), the density of \( V \).

**Assumption 30.** Let \( K(\cdot) \) denote a univariate kernel density. \( K(\cdot) \) is supported on \([-1,1]\), symmetric, and has order higher than \( \nu \) with bounded derivatives up to degree \( \nu \). \( K(1) = K(-1) = 0 \).

The key restrictions for \( \sqrt{n} \)-consistency of \( \hat{\alpha}_n \) are on the tail of \( V \) and \( \varepsilon \). Next, we introduce some definitions from the extreme value theory that help us characterizing the tail behaviors of probability distributions.

The cumulative distribution function (c.d.f.) \( F \) belongs to the domain of attraction of generalized extreme value distributions for its left tail if there exist sequences \( (\alpha_n)_{n \in \mathbb{N}}, (\beta_n)_{n \in \mathbb{N}} \), and \( G \), a c.d.f. indexed by a parameter \( \xi \), such that, for any independent draw \( (U_1, \ldots, U_n) \) from \( F \), \( \alpha_n(\min(U_1, \ldots, U_n) - \beta_n) \) converges in distribution to \( G \). In such a case, \( G \) belongs to the family of generalized extreme value distributions with a parameter \( \xi \) in which the parameter is called the extreme value (EV) index for the left tail. In particular, \( G \) is of type 1, 2, or 3 if for

- type 1 tails \( (\xi_t = 0) \): as \( z \to s_t \quad F(z + v\alpha) \sim F(z)e^v, \quad \forall v \in \mathbb{R} \),
- type 2 tails \( (\xi_t > 0) \): as \( z \to s_t = -\infty \quad F(vz) \sim v^{-1/\xi}F(z), \quad \forall v > 0 \),
- type 3 tails \( (\xi_t < 0) \): as \( z \to s_t > -\infty \quad F(vz) \sim v^{-1/\xi}F(z), \quad \forall v > 0 \).
in which \( a(z) := \int_{s_l}^{z} F(v)dv/F(z) \) for some \( z > s_l \) and \( s_l \) is the lower endpoint of the support of \( U \).

Similarly, \( F \) belongs to the domain of attraction of generalized extreme value distributions for its right tail if there exist sequences \((\alpha_n)_{n \in \mathbb{N}}, (\beta_n)_{n \in \mathbb{N}}\), and \( G \), a c.d.f. indexed by a parameter \( \xi \), such that, for any independent draw \((U_1, ..., U_n)\) from \( F \), 
\[ \alpha_n(\max(U_1, ..., U_n) - \beta_n) \]
converges in distribution to \( G \). In such a case, \( G \) belongs to the family of generalized extreme value distributions with a parameter \( \xi \) in which the parameter is called the extreme value (EV) index for the right tail. In particular, \( G \) is of type 1, 2, or 3 if for

1. **Type 1 tails** \((\xi_r = 0)\): as \( z \to s_u \)
   
   \[
   (1 - F)(z + va(z)) \sim (1 - F)(z)e^\nu, \quad \forall \nu \in \mathbb{R},
   
   \]

2. **Type 2 tails** \((\xi_r > 0)\): as \( z \to s_u = -\infty \)
   
   \[
   (1 - F)(vz) \sim v^{-1/\xi}(1 - F)(z), \quad \forall v > 0,
   
   \]

3. **Type 3 tails** \((\xi_r < 0)\): as \( z \to s_u > -\infty \)
   
   \[
   (1 - F)(vz) \sim v^{-1/\xi}(1 - F)(z), \quad \forall v > 0,
   
   \]

in which \( a(z) := \int_{z}^{s_u} (1 - F)(v)dv/(1 - F)(z) \) for some \( z < s_u \) and \( s_u \) is the upper endpoint of the support of \( U \).

In addition, we write \( G \in RV_\alpha(\infty) \) for a generic univariate function \( G \) if \( G \) is regularly varying with varying index \( \alpha \) at \( \infty \), i.e. \( \frac{G(xt)}{G(t)} \to x^\alpha \) as \( t \to \infty \) for any \( x > 0 \).

The inverse of a c.d.f \( G \) is written as \( G^\leftarrow(\tau) = \inf\{t : G(t) > \tau\} \). The inverse of a survival function \( 1 - G \) is \( (1 - G)^\leftarrow(\tau) = \inf\{t : (1 - G)(t) \leq \tau\} \).

Now we are ready to state the regularity conditions for the tails of \( V \) and \( \varepsilon \).

**Assumption 31.** Let \( F \) and \( F_\varepsilon \) be the c.d.f. of \( V \) and \( \varepsilon \), respectively.

1. \( F \) is in the attraction domain of EV distribution for both its left and right tails with EV indices \( \xi_r \) and \( \xi_l \), respectively.
2. \( \frac{\partial (1-F)^\leftarrow(\tau)}{\partial \tau} \in RV_{-\xi_r - 1}(0) \) and \( \frac{\partial F^\leftarrow(\tau)}{\partial \tau} \in RV_{-\xi_l - 1}(0) \).

\(^2\) Note here that \((1 - G)^\leftarrow(\tau) = G^\leftarrow(1 - \tau)\) if \( G \) is continuous at \((1 - G)^\leftarrow(\tau)\). Otherwise, \((1 - G)^\leftarrow(\tau)\) and \( G^\leftarrow(1 - \tau)\) are not necessarily the same. Throughout the paper, we consider the case in which the special regressor \( V \) and the error term \( \varepsilon \) are both continuous random variables.
(3) $F_\varepsilon$ is in the attraction domain of EV distribution for both its left and right tails with EV indices $\lambda_r$ and $\lambda_l$, respectively.

Assumption 31(1) and 31(2) are satisfied by almost all well-known continuous distributions and thus are mild. We refer readers to Dekkers and De Haan (1989) and Chernozhukov (2005c) for further discussion of these conditions. Assumption 31(3) is on the same footing as Assumption 31(1).

Next, we turn to the relative thickness of the tails of $V$ and $\varepsilon$, which has been identified by Khan and Tamer (2010) as the key condition for the convergence rate of semiparametric estimators of $\alpha$.

**Assumption 32.** For the right tail, one of the following three tail restrictions is satisfied and the symmetric conditions hold for the left tail.

(1) $\zeta_r > 0$ and $\lambda_r = 0$.

(2) $\zeta_r > 0$, $\lambda_r > 0$ and $\min(1 - \frac{1}{\nu}, \frac{1}{1 + \sigma}) > (1 + \zeta_r)^{\frac{\lambda_r}{\zeta_r(1 - \lambda_r)}}$.

(3) $\zeta_r = 0$, then $1 - F(t) = \exp(-T_r(t))$ with $T_r(t) \in RV_{d_1,r}(\infty)$, $1 - F_\varepsilon(t) = \exp(-D_r(t))$ with $D_r(t) \in RV_{d_1,r}(\infty)$, and $\infty \geq d_2,r > d_1,r \geq 0$.

Assumption 32 is the tail restriction for the existence of $\sqrt{n}$-consistent estimator of the intercept. Assumption 32(1) implies that if $\varepsilon$ has a rapidly varying tail and $V$ has regularly a varying tail, then the tail restrictions for $\sqrt{n}$-consistency of $\hat{\alpha}_n$ hold. Assumption 32(2) considers the case in which both $\varepsilon$ and $V$ have regularly varying tails. Then the tail restriction is on the relative magnitude of the two EV indices. Assumption 32(3) considers the case in which both tails are rapidly varying, in which the varying speed can no longer be described by its EV index. The condition presented here can be viewed as a restriction on the varying index of the logarithm of the tail.

Under the tail regularity conditions in Assumption 31, Assumption 32 is testable because the tails of both $V$ and $\varepsilon$ are identified. First, because $V_i$ is observable, the
tail of the c.d.f. of \( V \) can be identified from the data. In addition, because \( V \) is supported on the whole real line, we can identify the c.d.f. of \( \varepsilon - \alpha \) by

\[
F_\varepsilon(\alpha + v) = \mathbb{E}(Y|V = v), \quad \forall v \in \mathbb{R}
\]

and the parameter \( \alpha \). This implies that we can identify the c.d.f. of \( \varepsilon \).

Magnac and Maurin (2007) proposed a different set of tail restrictions in the same binary response model. Our tail restrictions are not nested by theirs. Conceptually, their tail restrictions are for point identification, whereas ours are for \( \sqrt{n} \)-consistency. The reason they have an identification problem is that they focus on the case in which the support of \( V \) is not the real line, so that the support of \( \varepsilon \) is not necessarily nested by the support of \( V \), especially when \( \varepsilon \) has full support \( \mathbb{R} \). On the other hand, we focus on the case in which \( V \) has full support \( \mathbb{R} \) so that \( \alpha \) is point identified. To derive the \( \sqrt{n} \)-consistency of their estimator, Magnac and Maurin (2007) in fact relied on the support of \( V \) being compact or on the high level assumptions used in Lewbel (2000). See, for example, their footnotes 9 and 10 on page 12. However, the high level assumptions in Lewbel (2000) may not hold in general because Khan and Tamer (2010) showed that \( \alpha \) has zero efficiency bound for some data generating processes (d.g.p.s) and thus the semiparametric efficiency bound is the worst-case bound. The restrictions provided in our paper are sufficient to rule out these d.g.p.s so that the worst-case bound becomes positive.

Now we can define our feasible trimming function \( \hat{I}_{n,i} \) and the two end points \( \hat{l}_n \) and \( \hat{r}_n \). Recall that \( F \) is the c.d.f. of the special regressor \( V \). The infeasible trimming we want to propose is \( I_{n,i} = 1\{V_i \in [l_n,r_n]\} \) where \( r_n = (1 - F)^{-\left(n^{-\rho_{r}}\right)} \) and \( l_n = F^{-\left(n^{-\rho_{l}}\right)} \) for some tuning parameter \( \rho_{r} \) and \( \rho_{l} \). Since \( F \) is unknown, \( l_n \) and \( r_n \) are infeasible. However, since \( r_n \) and \( l_n \) are the extremal quantile of special regressor \( V \), they can be estimated by order statistics. We denote the estimators for quantiles \( r_n \) and \( l_n \) by \( \hat{r}_n \) and \( \hat{l}_n \), in which \( \hat{r}_n = V_{(n-m_r+1)}^{(n)} \) and \( \hat{l}_n = V_{(m_l)}^{(n)} \). Here
Assumption 33. Let $h = c_h n^{-H}$ be the tuning parameter in the kernel density estimator, in which $c_h$ is a fixed positive constant.

1. If tail restriction Assumption 32(1) or 32(3) holds, then

$$\frac{1}{2\nu} < H < \min(1 - 2\rho_r(\zeta_r + 1), \frac{1}{2}).$$

2. If Assumption 32(2) holds, then

$$\frac{\lambda_r}{2\zeta_r(1 - \lambda_r)} < \rho_r < \frac{1}{2} - \frac{1}{2\nu} \left(\frac{1}{\zeta_r + 1}\right)$$

and

$$\frac{1}{2\nu} < H < \frac{1}{2} - (\zeta_r + 1)\rho_r.$$ 

3. Symmetric conditions for the left tail hold.

Tuning parameter $\rho_r$, $\rho_l$, and $H$ depend on EV indices $\zeta_r$ and $\zeta_l$ of $V$. Since $V$ is observable, we can consistently estimate $\zeta_r$ and $\zeta_l$. Then, under Assumption 32(1) or 32(3), the tuning parameter can be computed based on the estimator of $\zeta_r$ and $\zeta_l$. If instead, Assumption 32(2) holds, the choice of tuning parameter also depends on the EV indices $(\lambda_l, \lambda_r)$ of $\varepsilon$ that are identified. However, based on our limited knowledge, there are no estimators of $(\lambda_l, \lambda_r)$ directly available in the literature. To avoid estimating $(\lambda_l, \lambda_r)$, we can enhance Assumption 32(2) to hold with a known slack variable $\eta$, i.e.,

$$\left(\frac{1}{2} - \frac{1}{2\nu}\right) \frac{2\zeta_r}{\zeta_r + 1} \left(\frac{1}{\lambda_r} - 1\right) \geq 1 + \eta.$$
Then, \((\rho_r, H)\) with \(\rho_r \leq \frac{(1+\frac{\zeta_r}{\zeta_r+1})}{(1+\frac{1}{\zeta_r+1})}\) and \(\frac{1}{2^\nu} < H < \frac{1}{2} - (\zeta_r + 1)\rho_r\) satisfies Assumption 33 (2) and does not depend on the EV indices of \(\varepsilon\).

**Theorem 4.3.1.** Let \(Z_{n,i} = \frac{Y_i - \mathbb{I}(V_{i} > 0)}{f(V_i)} \hat{I}_{n,i}\). Under Assumptions 28 - 33,

\[
\begin{align*}
(1) & \quad \sqrt{n}(\hat{\alpha}_n - \alpha) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Z_{n,i} - \alpha) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Y_i - P_i}{f(V_i)} + o_p(1), \\
(2) & \quad \text{Let } \Sigma = E\left(\frac{Y_i - \mathbb{I}(V_{i} > 0)}{f(V_i)}\right)^2 - \alpha^2, \hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} Z_{n,i}^2 - \left(\frac{1}{n} \sum_{i=1}^{n} Z_{n,i}\right)^2, \text{ then } \Sigma < \infty \text{ and } \hat{\Sigma} \xrightarrow{p} \Sigma. \\
(3) & \quad \hat{\Sigma}^{-1/2} \sqrt{n}(\hat{\alpha}_n - \alpha) = \hat{\Sigma}^{-1/2} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Z_{n,i} - \alpha) \xrightarrow{d} N(0, 1).
\end{align*}
\]

Khan and Tamer (2010) pointed out that the reason for the non-existence of \(\sqrt{n}\)-consistent semiparametric estimator of \(\alpha\) is that \(E\left[\frac{Y_i - \mathbb{I}(V_{i} > 0)}{f(V_i)}\right]^2 = \infty\) for some d.g.p. Their intuition was that, when \(V\) has the full support \(\mathbb{R}, f(\cdot)\), the density of \(V\), vanishes at \(\pm \infty\). This induces the common “zero-denominator” problem. However, Theorem 4.3.1 shows that, under the tail restrictions, the d.g.p.s that correspond to the infinite second moment are ruled out. Here we still have the “zero-denominator” problem. But for this binary choice model, the numerator vanishes at \(V = \pm \infty\), too. If the numerator decays faster than the denominator at tails, \(\left[\frac{Y_i - \mathbb{I}(V_{i} > 0)}{f(V_i)}\right]^2\) is still integrable w.r.t. the density of \(V\). Our tail restrictions are sufficient for this purpose. For general inverse-propensity-weighted estimators, the numerator and denominator may not simultaneously vanish at the tails of some variable \(V\). That is one reason why we propose to trim \(V\), in contrast with Chaudhuri and Hill (2013), who tried to trim the whole estimand.
Notice that as sample size increases, the order statistic of \( V \) will diverge. This implies that the trimming interval will eventually become the real line. Thus even though the tail restrictions do not hold, the proposed estimator can still be consistent.

Theorem 4.3.1 extends the common knowledge that \( \sqrt{n} \)-consistency can be obtained in the binary response model if the special regressor has infinite variance. In fact, Assumption 32 allows for any moments of \( V \) to exist. This improvement is based on (1) additional knowledge of the tail behaviors of \( V \) and \( \varepsilon \), and (2) a careful calculation which is made possible by trimming based on extremal quantile.

Next, we show that the estimator \( \hat{\alpha}_n := \frac{\sum_{i=1}^n Z_{n,i}}{n} \) proposed in Theorem 4.3.1 is in fact asymptotically efficient and that the efficient function is \( \tilde{\psi} := \frac{Y - E(Y|V)}{f(V)} \).

**Corollary 4.3.1.** Under the conditions in Theorem 4.3.1, \( \tilde{\psi} \) is the efficient function for \( \alpha \) and \( \hat{\alpha}_n \) is asymptotically efficient with asymptotic distribution \( \mathcal{N}(0, E(\tilde{\psi}^2)) \).

Theorem 4.3.1 and Corollary 4.3.1 show that the tail restrictions in Assumption 32 are sufficient for the regular identification of \( \alpha \) when the support of \( \varepsilon \) is the whole real line and the proposed estimator is asymptotically efficient. In fact, the estimator is asymptotically equivalent to the estimator proposed in Lewbel (1997) when his high level assumptions for \( \sqrt{n} \)-consistency holds. The efficiency bound is also the same as the one derived in Magnac and Maurin (2007) and Jacho-Chávez (2009), although the underlying assumptions for \( \sqrt{n} \)-consistency are different. In fact, the tail restrictions in Assumption 32 do not affect the efficiency score. The same situation occurs in Magnac and Maurin (2007), as their tail symmetric condition for identification does not affect the efficiency bound.

Researchers may still be concerned about the necessity of this type of tail restriction. Assumption 34 characterizes situations of tails under which there does not exist any regular semiparametric estimator for \( \alpha \). It roughly means \( \alpha \) is not
√n-estimable. Assumption 34 can be viewed as the reverse situations of those considered in Assumption 32 in the sense that the roles of V and ε in Assumption 32 are reversed.

**Assumption 34.** $F_\varepsilon$, the c.d.f. of $\varepsilon$, is in the attraction domain of EV distribution for both its left and right tails with varying exponent $\lambda_r$ and $\lambda_l$. For the right tail, one of the following three tail restrictions is satisfied or symmetric conditions for the left tail.

1. $\zeta_r = 0$ and $\lambda_r > 0$
2. $\zeta_r > 0$, $\lambda_r > 0$ and $\lambda_r > \frac{\zeta_r}{2\gamma + 1}$.
3. $\zeta_r = 0$, $\lambda_r = 0$, then $1 - F(t) = \exp(-T_r(t))$ with $T_r(t) \in RV_{d_1,r}(\infty)$, $1 - F_\varepsilon(t) = \exp(-D_r(t))$ with $D_r(t) \in RV_{d_2,r}(\infty)$ and $\infty \geq d_{1,r} > d_{2,r} \geq 0$.

In short, Assumption 34 requires that the tails of V are thinner than the tails of $\varepsilon$. Section 3.1 of Khan and Tamer (2010) considers several examples of distributions of special regressor V and error term $\varepsilon$ and finds that when the tail of the special regressor V is as thin or thinner than the tail of the error term, the convergence rate for the estimator of the intercept term in (4.2.1) is slower than the parametric rate. Theorem 4.3.2 extends their specific observations to general situations when Assumption 34 holds.

**Theorem 4.3.2.** Under Assumptions 28-30 and 34, the asymptotic variance $E\left| \frac{\hat{\alpha} - \alpha}{f(V_\varepsilon)} \right|^2$ is infinite and there does not exist a regular estimator of $\alpha$.

Theorem 4.3.2 shows that for some d.g.p.s, the $\alpha$ is irregularly identified. This confirms the result in Khan and Tamer (2010) that, without tail restrictions, the semiparametric efficiency bound for $\alpha$ as the worst-case bound is zero. It also shows that the high-level assumptions for $\sqrt{n}$-consistency in Stoker (1991) do not hold in general.
The intuition for this theorem is the same as the one mentioned in Khan and Tamer (2010). $\frac{Y_i - E(Y_i|V_i)}{f(V_i)}$ is the efficient score when our tail restrictions hold in corollary 4.3.1. Suppose there exists an efficient estimator $\tilde{\alpha}_n$, which must have the following linear expansion:

$$\sqrt{n}(\tilde{\alpha}_n - \alpha) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Y_i - E(Y_i|V_i)}{f(V_i)} + o_p(1).$$

However, $E|\frac{Y_i - E(Y_i|V_i)}{f(V_i)}|^2 = E|\frac{Y_i - E(Y_i|V_i)}{f(V_i)}|^2 + \alpha^2 = \infty$ under Assumption 34. This implies that the second moment of $\frac{Y_i - E(Y_i|V_i)}{f(V_i)}$ is infinite and thus the efficient estimator of $\alpha$ has an infinite asymptotic variance. This is a contradiction.

4.4 Extensions

This section extends the method of asymptotic trimming by extremal quantiles to the estimator of binary choice models proposed by Dong and Lewbel (2015). In particular, we consider the following setup adapted from Corollary 1 of Dong and Lewbel (2015).

Assumption 35. Assume $Y = 1\{X'\beta + V \geq \varepsilon\}$, $E(Z\varepsilon) = 0$, $\Sigma_{xz} = EXZ'$ has full column rank, $EV = 0$, $V = S'\gamma + U$, $EU = 0$, $U \perp (S, \varepsilon)$ where $S = (X, Z)$. $U$ has density $f(U)$ and its support is the real line $\mathbb{R}$.

First, we note that $X$ is allowed to be endogenous since $EX\varepsilon \neq 0$. $Z$ is the instrumental variable used for identification. Second, we will assume that $U$ has a full support, which directly implies the support condition in Corollary 1 of Dong and Lewbel (2015). Last, since $U \perp (\varepsilon, S)$, we have $V \perp \varepsilon|S$, which is Assumption A.2 in Lewbel (2000). Then based on Theorem 1 of Dong and Lewbel (2015), $\beta$ is identified as

$$\beta = \Delta E\left(\frac{Z \frac{Y - 1\{V > 0\}}{f(V|S)}}{f(V|S)}\right), \quad (4.4.1)$$
in which \( \Delta = (\Sigma_{xz} W \Sigma_{xz}')^{-1} \Sigma_{xz} W \) and \( W \) is the usual weighting matrix. A popular choice for \( W \) is \( \Sigma_{zz}^{-1} \) where \( \Sigma_{zz} = E Z Z' \).

Until now, the condition that \( V = S' \gamma + U \) has not been used. The purpose of this condition is to reduce the dimensionality. We note that one undesirable feature of the estimator in (4.4.1) is that the kernel estimation of the conditional density \( f(V|S) \) suffers from the curse of dimensionality. It requires a relatively large dataset for the estimator to perform well. To overcome the curse of dimensionality, Dong and Lewbel (2015) imposed a parsimonious parametric model such that \( V = S' \gamma + U \). Here the linearity of the conditional expectation \( E(V|S) \) is not essential. The key is that \( U \) is additively separable and \( U \perp S \). Under this parametric assumption, we need to estimate only \( f(u) \), the density of \( U \), which is univariate. In fact, \( \beta \) can be identified as

\[
\beta = \Delta E \left( \frac{Z Y - 1\{V > 0\}}{f(U)} \right). \tag{4.4.2}
\]

As for the \( \sqrt{n} \)-consistency for the estimator of \( \beta \), the intuition from the previous section still applies: the convergence rate depends on the relative thickness of the tails of \( \varepsilon \) and \( U \). Next, we impose sufficient tail restrictions for the \( \sqrt{n} \)-consistency of an estimator of \( \beta \) that depends on a feasible asymptotic trimming scheme. Compared to the estimator of the intercept in the simplified model considered in Section 4.3, the additional difficulty here is that \( U_i \) is not directly observable. We propose to use the residual \( \hat{U}_i \) from the regression of \( V \) on \( S \) to substitute the unobservable \( U_i \). The feasible trimming function is based on the residual \( \hat{U}_i \).

In order to give a formal definition of our semiparametric estimator and asymptotic trimming function, we need the following assumption.

**Assumption 36.** The support of \( S \) is compact.

Assumption 36 implies that the tails of \( V \) and \( U \) are the same. This ensures that
the tail restrictions on $U$ are sufficient for $\sqrt{n}$-consistency. If the support of $S$ is in fact unbounded, we can trim $S$ by $\tau(S) = 1\{S \in S_0\}$ where $S_0$ is compact. Then $\beta$ can still be identified as
\[
\beta = \hat{\Delta}E\left( Z\tau(S)\frac{Y - 1\{V > 0\}}{f(U)} \right),
\]
in which $\hat{\Delta} = (\Sigma_{xz}W\Sigma'_{xz}W)^{-1}\Sigma_{xz}W$, $\Sigma_{xz} = EXZ'\tau(S)$, and $W$ is the usual weighting matrix. This implies that Assumption 36 holds without loss of generality.

Then we propose to estimate $\beta$ in two steps. In the first step, we regress $V$ on $S$ to obtain the OLS estimator $\hat{\gamma}$ of $\gamma$. Then we compute the residual $\hat{U}_i$ as
\[
\hat{U}_i = V_i - S'\hat{\gamma}.
\]
In the second step, $\beta$ is estimated as
\[
\hat{\beta} = (\Sigma_{xz}W_n\Sigma'_{xz})^{-1}\Sigma_{xz}W_n\hat{\Phi},
\]
in which $W_n$ is a (random) weighting matrix such that $W_n \overset{p}{\rightarrow} W$ for some positive definite (nonrandom) matrix $W$,
\[
\hat{\Sigma}_{xx} = \frac{1}{n}\sum_{i=1}^n Z_iX'_i,
\]
\[
\hat{\Phi} = \frac{1}{n}\sum_{i=1}^n Z_i \left( \frac{Y_i - 1\{V_i \geq 0\}}{\hat{f}(\hat{U}_i)} \right) \tilde{I}_{n,i}, \tag{4.4.3}
\]
\[
\hat{f}(\hat{U}_i) = \frac{1}{(n-1)h} \sum_{j \neq i} K\left( \frac{U_j - U_i}{h} \right), \quad \text{and} \quad \tilde{I}_{n,i} = 1\{\hat{U}_i \in (\tilde{l}_n, \tilde{r}_n)\}.
\]
Here $\tilde{r}_n = \hat{U}^{(n)}_{(n-m_r+1)}$ and $\tilde{l}_n = \hat{U}^{(n)}_{(m_l)}$ in which $m_r = \lfloor n^{1-\rho} \rfloor$ and $m_l = \lfloor n^{1-\rho} \rfloor$. $\tilde{I}_{n,i}$ is the feasible counterpart of $I_{n,i} = 1\{U_i \in S_n\}$, in which $S_n = (l_n, r_n)$, $r_n = (1 - F)^{\omega}(n^{-\rho})$, $l_n = F^{\omega}(n^{-\rho})$, and $F$ is the c.d.f. of $U$. 

129
**Assumption 37.** Let \( f(\cdot), F(\cdot) \) and \( F_\varepsilon(\cdot) \) denote the density of \( U \), the c.d.f. of \( U \), and the c.d.f. of \( \varepsilon \), respectively. Then Assumptions 29 and 31 hold for \( f(\cdot), F(\cdot), \) and \( F_\varepsilon(\cdot) \).

Since the tails of \( U \) and \( \varepsilon \) are the objects of interest, Assumption 37 relies on the same tail regularity assumptions that we used in the previous section. Next, we state our tail restrictions on the relative thickness of the tails between \( U \) and \( \varepsilon \).

**Assumption 38.** Recall that \( F \) and \( F_\varepsilon \) are the c.d.f. of \( U \) and \( \varepsilon \), respectively. Define \( \zeta_r \) and \( \zeta_l \) as the right and left EV index for \( U \) and \( \lambda_r \) and \( \lambda_l \) as the right and left EV index for \( \varepsilon \). For the right tail, one of the following three tail restrictions is satisfied, and the symmetric conditions hold for the left tail.

1. \( \zeta_r > 0 \) and \( \lambda_r = 0 \).
2. \( \zeta_r > 0, \lambda_r > 0, \) and \( \min(1 - \frac{4}{\nu}, \frac{1}{1+\sigma}) > \frac{(1+\zeta_l)\lambda_l}{\zeta_r(1-\lambda_r)} \).
3. \( \zeta_r = 0, 1 - F(t) = \exp(-T_r(t)) \) with \( T_r(t) \in RV_{d_1,r}(\infty) \), \( 1 - F_\varepsilon(t) = \exp(-D_r(t)) \) with \( D_r(t) \in RV_{d_2,r}(\infty) \), and \( \infty \geq d_2 > d_1 \geq 0 \).

Since \( \gamma \) is identified, so the tails of \( U_i = V_i - S_i'\gamma \). In addition, \( \beta \) is identified. Thus, as discussed after Assumption 32, the tails of \( \varepsilon \) are also identified. This implies that Assumption 38 is testable. It would be useful for future research to construct a feasible statistical test for the tail restrictions.

Given Assumption 38, we choose the two key tuning parameters \( \rho_r \) and \( \rho_l \) as follows:

**Assumption 39.** Let \( h = c_h n^{-H} \) be the tuning parameter in the kernel density estimator for some positive constants \( c_h \) and \( H \).

1. If Assumption 38(1) or (3) holds, then
   \[
   \frac{1}{2\nu} < H < \min\left(\frac{1}{4}(1 - 2\rho_r(\zeta_r + 1)), \frac{1}{8}\right).
   \]
(2) If Assumption 38(2) holds, then
\[ \frac{\lambda_r}{2\zeta_r(1 - \lambda_r)} < \rho_r < \left(1 - \frac{1}{2\nu}\right) \frac{1}{(\zeta_r + 1)} \]
and
\[ \frac{1}{2\nu} < H < \frac{1}{8}(1 - 2(\zeta_r + 1)\rho_r). \]

(3) Symmetric conditions for the left tail hold.

Comparing Assumption 39 with Assumption 33, the key difference is the upper bound of $H$. This is due to the fact that $U_i$ is not observed and the estimator $\hat{U}_i$ is used to compute $\hat{f}(\cdot)$, the kernel estimator of the density of $U$. If $U_i$ is observed, the accuracy of the univariate kernel density estimator is $O_p(\sqrt{n})$. Now since $\hat{U}_i$ is used, the accuracy becomes $O_p(\frac{1}{\sqrt{nh}})$.

The next theorem establishes the $\sqrt{n}$-consistency of $\hat{\beta}$.

**Theorem 4.4.1.** If $W_n \xrightarrow{p} W$ for some positive definite (nonrandom) matrix $W$, and if Assumptions 30-, and 35 - 39 hold, then
\[ E|\Psi_i|^{2+\sigma} < \infty \]
and
\[ \sqrt{n}(\hat{\beta} - \beta) = \left(\Sigma_{xz}W\Sigma_{xz}^t\right)^{-1}\Sigma_{xz}W \frac{1}{\sqrt{n}} \sum_{i=1}^n (\Psi_i - (Z_iX_i' - \Sigma_{xz})\beta) + o_p(1) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\beta}), \]
in which
\[ \Psi_i = \frac{Z_i(Y_i - 1\{V_i > 0\})}{f(U_i)} + \frac{E(Z_i(Y_i - 1\{V_i > 0\})|U_i)f'(U_i)(S_i - ES_i)'\Sigma_{ss}^{-1}S_i}{f^2(U_i)} \]
\[ \Sigma_\beta = (\Sigma' W \Sigma W')^{-1} \Sigma' W \Sigma S \Sigma + (\Sigma' W \Sigma W')^{-1}, \]

\[ \Sigma_0 = E(\Psi_i - (Z_i X_i' - \Sigma) \beta)(\Psi_i - (Z_i X_i' - \Sigma) \beta)' , \]

and \( \Sigma_{ss} = ESS' \).

4.5 Simulations

In this section, we check the finite sample performance of our estimator with asymptotic trimming using extremal quantiles. We use the same model investigated by the simulation section of Lewbel (1997):

\[ Y_i = 1\{-1 + V_i - \varepsilon_i > 0\}. \]

We consider six simulation designs. The first five satisfy the tail restrictions formulated in Assumption 32 so that the proposed estimator is \( \sqrt{n} \)-consistent and asymptotically normal. The last design satisfies the tail restrictions formulated in Assumption 34 so that there is no regular estimators exist. All simulations are conducted for 1,000 replications and the sample sizes considered are \( 200, 400, 800, 1600, 3200, 6400 \).

For each design, we report the bias and root-mean-square error of our estimator and the estimator suggested in Lewbel (1997). For the latter, Lewbel (1997) suggested trimming out the observations for which the estimated density is too small. In particular, he used

\[ \hat{\alpha}_L(b) = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i - 1\{V_i > 0\}}{f(V_i)} 1\{\hat{f}(V_i) > b\} \]

for some small but fixed constant \( b \). In simulations, we consider three choices of \( b \), namely \( b \in (0.05, 0.01, 0.002) \).

To estimate the density, we use the fourth-order Epanechnikov kernel,\(^3\) that is,

\[ k(u) = \frac{45}{32} (1 - \frac{7}{3} u^2) (1 - u^2) 1\{|u| < 1\}. \]

\(^3\) This implies \( \nu = 4 \).
The tuning parameter for the kernel density estimation takes the form of \( h := c_h n^{-H} \) where \( c_h \) is a positive constant and \( H \) is defined in Assumption 33. Since the bias is of order of \( h^4 \), based on Powell and Stoker (1996) and the results of numerical integration in Lewbel (1997), we set the constant as

\[
c_h = \left( \frac{2.532 \times 8}{0.0204 \times 2 \times s} \right)^{\frac{1}{8+s}},
\]

in which \( s \) is the dimension of \( V \) and \( s = 1 \) in this model.

For \( H \), when computing \( \hat{\alpha}_L(b) \) with \( b \in (0.05, 0.01, 0.002) \), we use the optimal rate \( H^* = \frac{2}{8+s} \). Lewbel (1997) also used the optimal rate \( \frac{2}{4+s} \). The two optimal rates are different because in Lewbel (1997)’s simulation, a second-order kernel is used, whereas here, we use a fourth-order kernel.

When computing \( \hat{\alpha}_n \), we will choose the optimal rate \( H^* \) if possible. However, the optimal rate may not satisfy Assumption 33 in some designs. Next, we will specify the rate based on each case.

**Design 1**

\( V \sim T(6) \) and \( \varepsilon \sim \mathcal{N}(0, 1) \). The EV indices for \( V \) are \( \xi_l = \xi_r = \frac{1}{6} \) and \( \lambda_l = \lambda_r = 0 \) for \( \varepsilon \). This implies that Assumption 31(1) holds. By choosing \( \rho_r = \rho_l = \frac{1}{4} \), Assumption 33(1) implies that

\[
\frac{1}{8} < H < \frac{5}{12}.
\]

Since the optimal rate \( H^* = \frac{2}{9} \) satisfies the above inequality, we choose \( H = H^* \) when computing \( \hat{\alpha}_n \). The next table shows the bias and root-mean-square error of the estimators. “Extremal” represents the new estimator \( \hat{\alpha}_n \) proposed in this paper. L1–L3 represent the estimators \( \hat{\alpha}_L(b) \) for \( b = 0.05, 0.01, 0.002 \), respectively.
Table 4.1: Bias and root-MSE

<table>
<thead>
<tr>
<th>N</th>
<th>Extremal</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
<th>Extremal</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.050</td>
<td>0.071</td>
<td>0.022</td>
<td>0.020</td>
<td>0.142</td>
<td>0.143</td>
<td>0.153</td>
<td>0.154</td>
</tr>
<tr>
<td>400</td>
<td>0.022</td>
<td>0.067</td>
<td>0.012</td>
<td>0.009</td>
<td>0.103</td>
<td>0.110</td>
<td>0.109</td>
<td>0.112</td>
</tr>
<tr>
<td>800</td>
<td>0.011</td>
<td>0.063</td>
<td>0.010</td>
<td>0.007</td>
<td>0.080</td>
<td>0.090</td>
<td>0.079</td>
<td>0.081</td>
</tr>
<tr>
<td>1,600</td>
<td>0.004</td>
<td>0.062</td>
<td>0.007</td>
<td>0.003</td>
<td>0.058</td>
<td>0.077</td>
<td>0.056</td>
<td>0.060</td>
</tr>
<tr>
<td>3,200</td>
<td>0.002</td>
<td>0.060</td>
<td>0.005</td>
<td>0.002</td>
<td>0.042</td>
<td>0.069</td>
<td>0.041</td>
<td>0.042</td>
</tr>
<tr>
<td>6,400</td>
<td>0.002</td>
<td>0.061</td>
<td>0.005</td>
<td>0.002</td>
<td>0.029</td>
<td>0.065</td>
<td>0.028</td>
<td>0.029</td>
</tr>
</tbody>
</table>

Since the root-mean-square errors for $\hat{\alpha}_n$ decrease at rate $\sqrt{2}$ as sample size doubles, $\hat{\alpha}_n$ is $\sqrt{n}$-consistent. This provides evidence that even $V$ has a finite second moment, $\alpha$ is still $\sqrt{n}$-estimable. In addition, when $b = 0.05$, the estimator L1 has a non-vanishing bias. This is not surprising because the threshold $b$ does not vary as sample size increases. When we choose a smaller $b$, the bias for estimators L2 and L3 is smaller. But they are still fixed. So asymptotically, the bias cannot vanish, although for the current sample sizes considered, the bias is relatively small compared to the estimation error. Last, our estimator “Extremal” is asymptotically valid and performs as well as L2 and L3 in terms of root-mean-square error. In fact, we have shown that $\hat{\alpha}_n$, the “Extrmeal” estimator, is efficient.

**Design 2**

Next, we consider the case in which $\varepsilon$ is not symmetrically distributed. In particular, we set $V \sim T(6)$ and $\varepsilon = \frac{e_1 + e_2 + e_3 - 2}{\sqrt{3}}$ where $(e_1, e_2, e_3)$ are independent standard normals. As in the first design, $\xi_l = \xi_r = \frac{1}{6}$ and $\lambda_l = \lambda_r = 0$. This implies that Assumption 32(1) holds and $\hat{\alpha}_n$ is $\sqrt{n}$-consistent. We choose the same set of tuning parameters as in Design 1; that is,

$$\rho_r = \rho_l = \frac{1}{4}, \quad H = H^*.$$
Table 4.2: Bias and root-MSE

<table>
<thead>
<tr>
<th>N</th>
<th>Extremal</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
<th>Extremal</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.089</td>
<td>0.106</td>
<td>0.039</td>
<td>0.038</td>
<td>0.150</td>
<td>0.154</td>
<td>0.162</td>
<td>0.163</td>
</tr>
<tr>
<td>400</td>
<td>0.057</td>
<td>0.104</td>
<td>0.037</td>
<td>0.029</td>
<td>0.115</td>
<td>0.133</td>
<td>0.119</td>
<td>0.126</td>
</tr>
<tr>
<td>800</td>
<td>0.038</td>
<td>0.101</td>
<td>0.032</td>
<td>0.019</td>
<td>0.088</td>
<td>0.116</td>
<td>0.089</td>
<td>0.101</td>
</tr>
<tr>
<td>1,600</td>
<td>0.020</td>
<td>0.099</td>
<td>0.028</td>
<td>0.007</td>
<td>0.069</td>
<td>0.109</td>
<td>0.067</td>
<td>0.080</td>
</tr>
<tr>
<td>3,200</td>
<td>0.014</td>
<td>0.100</td>
<td>0.030</td>
<td>0.010</td>
<td>0.051</td>
<td>0.105</td>
<td>0.050</td>
<td>0.054</td>
</tr>
<tr>
<td>6,400</td>
<td>0.009</td>
<td>0.101</td>
<td>0.030</td>
<td>0.009</td>
<td>0.040</td>
<td>0.103</td>
<td>0.041</td>
<td>0.040</td>
</tr>
</tbody>
</table>

We see that \( \hat{\alpha}_n \) is indeed \( \sqrt{n} \)-consistent. In addition to L1, now the bias for L2 also does not vanish, and the bias for “Extremal” and L3 are still small. This again indicates that the biases for L1–L3 does not vanish and thus \( \hat{\alpha}_L(b) \) is not asymptotically valid.

**Design 3**

Here we consider the case in which \( V \) is not symmetrically distributed and the tails of \( \varepsilon \) decay polynomially. In particular,

\[
V = T_1 + T_2^2 - 2, \quad V \sim T(11)
\]

where \( T_1 \sim T(6) \), \( T_2 \sim T(4) \), and \( T_1 \perp T_2 \). We have \( \xi_r = \frac{1}{2}, \xi_l = \frac{1}{6} \), and \( \lambda_r = \lambda_l = \frac{1}{11} \).

It is easy to check that Assumption 32(2) holds. Therefore, \( \hat{\alpha}_n \) is \( \sqrt{n} \)-consistent. By Assumption 33(2), \( \rho_l \) and \( \rho_r \) need to satisfy

\[
\frac{1}{10} < \rho_r < \frac{1}{4}, \quad \frac{42}{140} < \rho_l < \frac{45}{140}.
\]

So we choose \( \rho_r = \frac{1}{5} \) and \( \rho_l = \frac{44}{140} \). This further implies that

\[
\frac{1}{8} < H < \frac{1}{5}, \quad \frac{1}{8} < H < \frac{2}{15}.
\]

\( H = \frac{31}{240} \) suffices.
From the above table, we note that \( \hat{\alpha}_n \) is \( \sqrt{n} \)-consistent and our method can handle the situation in which the tail behaviors of \( V \) are asymmetric. Although the tuning parameter \( H \) is not optimal, the root-mean-square error of our estimator "Extremal" is very close to L1–L3. This is consistent with the theoretical result that \( \hat{\alpha}_n \) is efficient.

**Design 4**

For the fourth design, we consider the case in which \( V \) has exponentially decaying tails. In particular, \( V \sim \mathcal{N}(0,1) \), \( \varepsilon = \text{sign}(e_1)|e_1|^{1/2} \), and \( e_1 \sim \mathcal{N}(0,1) \). In this case, \( \xi_r = \xi_l = \lambda_r = \lambda_l = 0 \), but Assumption 32(3) holds because \( d_{1,r} = d_{1,l} = 2 \) and \( d_{2,r} = d_{2,l} = 6 \). By Assumption 33(1), we can set \( \rho_r = \rho_l = \frac{1}{4} \) and \( H = H^* \).

<table>
<thead>
<tr>
<th>N</th>
<th>Extremal</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
<th>Extremal</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.029</td>
<td>0.065</td>
<td>0.048</td>
<td>0.046</td>
<td>0.234</td>
<td>0.227</td>
<td>0.231</td>
<td>0.234</td>
</tr>
<tr>
<td>400</td>
<td>0.019</td>
<td>0.043</td>
<td>0.027</td>
<td>0.025</td>
<td>0.168</td>
<td>0.163</td>
<td>0.166</td>
<td>0.169</td>
</tr>
<tr>
<td>800</td>
<td>0.018</td>
<td>0.031</td>
<td>0.019</td>
<td>0.018</td>
<td>0.119</td>
<td>0.117</td>
<td>0.117</td>
<td>0.119</td>
</tr>
<tr>
<td>1,600</td>
<td>0.011</td>
<td>0.020</td>
<td>0.011</td>
<td>0.010</td>
<td>0.081</td>
<td>0.080</td>
<td>0.081</td>
<td>0.081</td>
</tr>
<tr>
<td>3,200</td>
<td>0.007</td>
<td>0.012</td>
<td>0.004</td>
<td>0.004</td>
<td>0.054</td>
<td>0.054</td>
<td>0.053</td>
<td>0.053</td>
</tr>
<tr>
<td>6,400</td>
<td>0.005</td>
<td>0.010</td>
<td>0.003</td>
<td>0.002</td>
<td>0.041</td>
<td>0.040</td>
<td>0.040</td>
<td>0.040</td>
</tr>
</tbody>
</table>

From the table, we see that even though every moment of \( V \) exists, our estimator \( \hat{\alpha}_n \) can still be \( \sqrt{n} \)-consistent.
Design 5

In this design, we consider the case in which both $V$ and $\varepsilon$ have exponentially decaying tails and the c.d.f. of $\varepsilon$ is not symmetric. In particular, $V = c_1^3, \varepsilon = \frac{e_2 + e_3^2 + e_4^2 - 2}{\sqrt{3}}$ where $e_1, e_2, e_3$, and $e_4$ are standard normally distributed and mutually independent. In this case, $\xi_r = \xi_l = \lambda_r = \lambda_l = 0$, but Assumption 32(3) holds. We can choose $\rho_r, \rho_l$, and $H$ as we did in Design 4; that is $\rho_r = \rho_l = \frac{1}{4}$ and $H = H^*$. 

<table>
<thead>
<tr>
<th>N</th>
<th>Extremal</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
<th>Extremal</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.081</td>
<td>0.250</td>
<td>0.082</td>
<td>0.080</td>
<td>0.219</td>
<td>0.290</td>
<td>0.218</td>
<td>0.219</td>
</tr>
<tr>
<td>400</td>
<td>0.048</td>
<td>0.232</td>
<td>0.057</td>
<td>0.048</td>
<td>0.163</td>
<td>0.256</td>
<td>0.157</td>
<td>0.163</td>
</tr>
<tr>
<td>800</td>
<td>0.042</td>
<td>0.231</td>
<td>0.049</td>
<td>0.042</td>
<td>0.122</td>
<td>0.245</td>
<td>0.117</td>
<td>0.122</td>
</tr>
<tr>
<td>1,600</td>
<td>0.032</td>
<td>0.223</td>
<td>0.038</td>
<td>0.032</td>
<td>0.084</td>
<td>0.231</td>
<td>0.083</td>
<td>0.084</td>
</tr>
<tr>
<td>3,200</td>
<td>0.026</td>
<td>0.222</td>
<td>0.031</td>
<td>0.027</td>
<td>0.063</td>
<td>0.226</td>
<td>0.063</td>
<td>0.063</td>
</tr>
<tr>
<td>6,400</td>
<td>0.021</td>
<td>0.215</td>
<td>0.025</td>
<td>0.021</td>
<td>0.045</td>
<td>0.218</td>
<td>0.046</td>
<td>0.045</td>
</tr>
</tbody>
</table>

From the table, we see that when $b$ is relatively large, as in L1, the bias does not vanish. When $b$ is small, the estimator proposed in this paper has a similar performance as L2 and L3, and is $\sqrt{n}$-consistent.

Design 6

Last, we consider the case in which Assumption 32 does not hold. In particular, $V \sim T(6)$ and $\varepsilon \sim T(2)$. This implies that $\xi_r = \xi_l = \frac{1}{6}, \lambda_r = \lambda_l = \frac{1}{2}$, and Assumption 34(2) holds. Then based on Theorem 4.3.2, there does not exist any regular estimator for $\alpha$. When computing our estimator, we choose $\rho_r = \rho_l = \frac{1}{4}$ and $H = H^*$. 

137
Table 4.6: Bias and root-MSE

<table>
<thead>
<tr>
<th>N</th>
<th>Extremal</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
<th>Extremal</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.170</td>
<td>0.188</td>
<td>0.088</td>
<td>0.085</td>
<td>0.241</td>
<td>0.242</td>
<td>0.253</td>
<td>0.256</td>
</tr>
<tr>
<td>400</td>
<td>0.128</td>
<td>0.184</td>
<td>0.096</td>
<td>0.072</td>
<td>0.191</td>
<td>0.212</td>
<td>0.189</td>
<td>0.219</td>
</tr>
<tr>
<td>800</td>
<td>0.102</td>
<td>0.185</td>
<td>0.092</td>
<td>0.057</td>
<td>0.158</td>
<td>0.200</td>
<td>0.151</td>
<td>0.184</td>
</tr>
<tr>
<td>1,600</td>
<td>0.078</td>
<td>0.182</td>
<td>0.089</td>
<td>0.047</td>
<td>0.131</td>
<td>0.190</td>
<td>0.122</td>
<td>0.157</td>
</tr>
<tr>
<td>3,200</td>
<td>0.062</td>
<td>0.182</td>
<td>0.089</td>
<td>0.051</td>
<td>0.107</td>
<td>0.186</td>
<td>0.107</td>
<td>0.110</td>
</tr>
<tr>
<td>6,400</td>
<td>0.046</td>
<td>0.181</td>
<td>0.088</td>
<td>0.050</td>
<td>0.089</td>
<td>0.183</td>
<td>0.099</td>
<td>0.085</td>
</tr>
</tbody>
</table>

From the table, we first see that no estimator is $\sqrt{n}$-consistent. Second, the bias for our estimator “Extremal” is still decreasing, while the bias for L1 and L2 do not vanish.

Summary
In general, we obtain three notable findings from this simulation study. First, when Assumption 32 holds, our estimator is $\sqrt{n}$-consistent. In this case, Lewbel (1997)’s estimator based on trimming out observations with small density has non-vanishing and dominant bias when the trimming threshold is not small enough. When the threshold is small enough, the bias of Lewbel (1997)’s estimator is not dominant given the sample size we considered, and its finite sample performances are similar to those of our estimator. Second, we verify that even if $V$ has all its moments exist, the existence of a $\sqrt{n}$-consistency estimator of $\alpha$ is still possible. Last, when Assumption 34 holds, neither our estimator nor Lewbel (1997)’s is $\sqrt{n}$-consistent. However, our estimator is still consistent because the bias is vanishing.

4.6 Conclusion

Because the intercept of the binary response model is irregularly identified, the convergence rate for its semiparametric estimator depends on the tail behaviors of the special regressor $V$ and the unobservable $\varepsilon$. This paper proposes a set of primitive
tail restrictions that guarantee the existence of a $\sqrt{n}$-consistent estimator for the intercept. In addition, we also proposed another set of primitive tail restrictions under which there does not exist any regular estimator for the intercept. Given the tail restrictions for $\sqrt{n}$-consistency, we propose a semiparametric estimator for the intercept by trimming based on extremal quantiles of the special regressor and show that the estimator is efficient. Last, we extend the method of trimming by extremal quantiles to allow for endogenous covariates $X$. 
Appendix A

Appendix for Chapter 2

A.1 Appendix

A.1.1 Illustrative examples

In this section, I consider four different types of conditional boundaries of $Y_1$ given $X$: single minimizer, multiple minimizers, continuum minimizers and mixture minimizers. For each of the boundary behavior, I compute the limiting objective function based on the theoretical results in Section 2.4. The results derived in this section are further used as the baseline models in the simulation study.

Example 1 (Single minimizer):

$$Y_1 = 0.5 + (X - 0.2)^2 + \varepsilon, \quad D = 1\{\eta \leq P(x)\}, \quad P(x) = 0.25 + x^2/2,$$

in which $X \sim \text{Uniform}[0, 1], \ \varepsilon \sim \text{Beta}(1, 2), \ \eta \sim \text{Uniform}[0, 1], \ X, \varepsilon, \eta$ are independent.

In this example, $\varpi(x)$, the conditional boundary of $Y$, is equal to $0.5 + (X - 0.2)^2$ and has a unique minimizer at $x = 0.2$. In addition, the EV index for $Y$ is $-1/1.5$. In general, the EV index is $-1/(\alpha + 0.5)$ where $\alpha$ is the first parameter of the Beta
distribution. Hence by Corollary 2.4.2, sequence \((D_i, \mathcal{E}_i)\) is i.i.d, \(D_i\) is Bernoulli distributed with success probability \(P(0.2)\), \(\mathcal{E}_i \perp D_i\), \(J_i = (\sum_{i=1}^n \mathcal{E}_i)^{1/5}\) in which \(\mathcal{E}_i\) is standard exponentially distributed, and

\[
\hat{Z}_{1,n}(k) \approx Z_{1,\infty}(k) := \arg \min_{z \in \mathbb{R}} -kz + \sum_i \frac{D_i}{P(0.2)}l_\delta(J_i, z).
\]

Example 2: (Multiple minimizers)

\[Y_1 = 0.5 + (|X - 0.3| - 0.1)^2 + \varepsilon, \quad D = \mathbb{1}\{\eta \leq P(x)\}, \quad P(x) = 0.25 + x^2/2,\]

in which \(X \sim \text{Uniform}[0, 1], \varepsilon \sim \text{Beta}(1, 2), \eta \sim \text{Uniform}[0, 1], X, \varepsilon, \eta\) are independent.

In this example, \(\pi(x)\), the conditional boundary of \(Y_i\) is \(0.5 + (|X - 0.3| - 0.1)^2\) and has two minimizers \(x_1 = 0.2\) and \(x_2 = 0.4\). It is easy to compute that \(S_{y,1} = [0.2 - \sqrt{y-0.5}, 0.2 + \sqrt{y-0.5}]\), \(S_{y,2} = [0.4 - \sqrt{y-0.5}, 0.4 + \sqrt{y-0.5}]\) and \(P_1 = P_2 = 1/2\). Again, the EV index for \(Y\) is \(-1/1.5\). In general, the EV index is \(-1/(\alpha + 0.5)\) where \(\alpha\) is the first parameter of the Beta distribution. Hence by Corollary 2.4.2, sequence \((D_i, \mathcal{X}_i, \mathcal{E}_i)\) is i.i.d, \(D_i\) is Bernoulli distributed with success probability \(P(\mathcal{X}_i)\) conditional on \(\mathcal{X}_i\), \(\mathcal{X}_i\) is equal to \(x_1 = 0.2\) or \(x_2 = 0.4\) with equal probability, \(\mathcal{E}_i \perp (\mathcal{X}_i, D_i)\), \(J_i = (\sum_{i=1}^n \mathcal{E}_i)^{1/5}\) where \(\mathcal{E}_i\) is standard exponentially distributed, and

\[
\hat{Z}_{1,n}(k) \approx Z_{1,\infty}(k) := \arg \min_{z \in \mathbb{R}} -kz + \sum_i \frac{D_i}{P(\mathcal{X}_i)}l_\delta(J_i, z).
\]

Example 3: (Continuum minimizers)

\[Y_1 = 0.5 + (X + 0.5)\varepsilon, \quad D = \mathbb{1}\{\eta \leq P(x)\}, \quad P(x) = 0.25 + x^2/2,\]

in which \(X \sim \text{Uniform}[0, 1], \varepsilon \sim \text{Beta}(1, 2), \eta \sim \text{Uniform}[0, 1], X, \varepsilon, \eta\) are independent.
In this example, \( \varpi(x) \), the conditional boundary of \( Y \) is flat. It is easy to compute that the EV index of \( Y \) is \(-1\) (\(-1/\alpha\) in general where \( \alpha \) is the first parameter of the Beta distribution). Hence by Corollary 2.4.3, sequence \((D_i, X_i, \mathcal{E}_i)\) is i.i.d, \( D_i \) is Bernoulli distributed with success probability \( P(X_i) \) conditional on \( X_i \), \( X_i \) is continuously distributed over \([0, 1]\) with density \( x + 0.5 \). In general, the density is

\[
\frac{\left(\frac{1}{\alpha} + 1\right)(x + 0.5)^{1/\alpha}}{1.5^{1/\alpha + 1} - 0.5^{1/\alpha + 1}}
\]

where \( \alpha \) is the first parameter of the Beta distribution. \( \mathcal{E}_i \perp (D_i, X_i) \), \( \mathcal{J}_i = \sum_{i=1}^{\infty} \mathcal{E}_i \) where \( \mathcal{E}_i \) is standard exponentially distributed, and

\[
\hat{Z}_{1,n}(k) \rightarrow Z_{1,\infty}(k) := \arg \min_{x \in \mathbb{R}} -kz + \sum_{i} \frac{D_i}{P(X_i)} l_i(\mathcal{J}_i, z).
\]

Example 4: (Mixture minimizers)

\( Y_1 = 0.5 + (|X - 0.3| - 0.1)^2 1\{X \in [0, 0.6]\} + (1\{X > 0.5\} - 1\{X \in [0.7, 0.8]\}) + (X + 0.5)\varepsilon, \)

\( D = 1\{\eta \leq P(x)\}, \quad P(x) = 0.25 + x^2/2, \)

where \( X \) takes value 0.2 with probability 0.1, 0.4 with probability 0.1 and is uniformly distributed on \([0.5, 1]\). \( \varepsilon \sim \text{Beta}(1, 2), \quad \eta \sim \text{Uniform}[0, 1], \quad X, \varepsilon, \eta \) are independent.

In this example, \( \varpi(x) \), the conditional boundary of \( Y \), is

\[
(\max\{|X - 0.3| - 0.1\} 1\{X \in [0, 0.6]\} + (1\{X > 0.5\} - 1\{X \in [0.7, 0.8]\})).
\]

\( \varpi(x) \) achieves its minimum at \( x_1 = 0.2, \ x_2 = 0.4 \) and \( x \in [0.7, 0.8] \). It is easy to compute that \( P_1^d = 1/3.6, \ P_2^d = 1/3.6, \ P_1^c = 1.6/3.6 \). Further more, the EV index for \( Y \) is \(-1\). In general, the EV index is \(-1/\alpha\), where \( \alpha \) is the first parameter of the Beta distribution. Hence by Corollary 2.4.4, sequence \((D_i, X_i, \mathcal{E}_i)\) is i.i.d, \( D_i \) is Bernoulli distributed with success probability \( P(X_i) \) conditional on \( X_i \), \( X_i \) is a
mixture distribution which has mass 1/3.6 at point 0.2, mass 1/3.6 at point 0.4, and is continuously distributed on [0.7, 0.8] with density \( \frac{32}{9} (x + 0.5) \). In general, the density is

\[
\frac{4\left(\frac{1}{\alpha} + 1\right)(x + 0.5)^{1/\alpha}}{9(1.3^{1/\alpha} + 1.2^{1/\alpha})}
\]

where \( \alpha \) is the first parameter of the Beta distribution. \( \mathcal{E}_i \perp (X_i, D_i), J_i = \sum_{i=1}^n \mathcal{E}_i \) where \( \mathcal{E}_i \) is standard exponentially distributed, and

\[
\hat{Z}_{1,n}(k) \sim Z_{1,\infty}(k) := \arg\min_{z \in \mathbb{R}} -kz + \sum_i \frac{D_i}{P(X_i)} I_{D_i}(J_i, z).
\]

**A.1.2 Details of simulation designs**

For all d.g.p’s, the error term \( \varepsilon_1 \) is generated from a Beta distribution with parameter (1, 2) and \( \varepsilon_0 \) is generated from a Beta distribution with parameter (1.5, 2). They are independent of each other as well as covariate \( X \). The treatment status \( D = 1 \{ U \leq P(x) \} \) where \( U \) is a uniformly distributed random variable independent of \( (\varepsilon_1, \varepsilon_0, X) \) and \( P(x) \) is the propensity score that takes the form of \( 0.25 + 0.5x^2 \). The potential outcomes (\( Y_1, Y_0 \)) are generated based on one of the following four models. For \( j = 0, 1, \)

1. Model \( (A_j) \):

   \[
   Y_j = a_{1,j} + (X - a_{2,j})^2 + \varepsilon_j,
   \]

   \( X \) is uniformly distributed on \([0, 1]\), \((a_{1,1}, a_{2,1}) = (0.5, 0.2)\), and \((a_{1,0}, a_{2,0}) = (0.2, 0.3)\).

2. Model \( (B_j) \):

   \[
   Y_j = b_{1,j} + (|X - b_{2,j}| - b_{3,j})^2 + \varepsilon_j,
   \]

   \( X \) is uniformly distributed on \([0, 1]\), \((b_{1,1}, b_{2,1}, b_{3,1}) = (0.5, 0.3, 0.1)\), and \((b_{1,0}, b_{2,0}, b_{3,0}) = (0.3, 0.2, 0.15)\).
3. Model \((C_j)\):

\[ Y_j = c_{1,j} + (X + c_{2,j})\varepsilon_j, \]

\(X\) is uniformly distributed on \([0, 1]\), \((c_{1,1}, c_{2,1}) = (0.5, 0.5)\), and \((c_{1,0}, c_{2,0}) = (0.3, 0.2)\).

4. Model \((D_j)\):

\[ Y_j = d_{1,j} + (|X-d_{2,j}|-d_{3,j})^2 1\{X < 0.6\} + (1\{X > 0.5\} - 1\{0.7 < X < 0.8\}) + (X+0.5)\varepsilon_j, \]

\(X\) takes values 0.2 or 0.4 with 0.1 probability and is uniform over \([0.5, 1]\), \((d_{1,1}, d_{2,1}, d_{3,1}) = (0.5, 0.3, 0.1)\), and \((d_{1,0}, d_{2,0}, d_{3,0}) = (0.3, 0.3, 0.1)\).

The 16 simulation designs considered in Section 4.5 can be summarized in the following table where the first coordinate represents \(Y_1\) and the second coordinate represents \(Y_0\).

<table>
<thead>
<tr>
<th>((A_1, A_0))</th>
<th>((A_1, B_0))</th>
<th>((A_1, C_0))</th>
<th>((A_1, D_0))</th>
</tr>
</thead>
<tbody>
<tr>
<td>((B_1, A_0))</td>
<td>((B_1, B_0))</td>
<td>((B_1, C_0))</td>
<td>((B_1, D_0))</td>
</tr>
<tr>
<td>((C_1, A_0))</td>
<td>((C_1, B_0))</td>
<td>((C_1, C_0))</td>
<td>((C_1, D_0))</td>
</tr>
<tr>
<td>((D_1, A_0))</td>
<td>((D_1, B_0))</td>
<td>((D_1, C_0))</td>
<td>((D_1, D_0))</td>
</tr>
</tbody>
</table>

**A.1.3 Additional simulation results**

*Simulation results with \(n = 300\)*

Tables A.2 and A.3 report the coverage of BN-CI and NN-CI as well as their corresponding median lengths. I am interested in the QTE at quantile order \(k = (5, 10, 20, 40)\). In this case, the corresponding quantile indices are

\[ \tau_n = (0.017, 0.033, 0.067, 0.133). \]

\(Y_1\) and \(Y_0\) have four different conditional boundary structures: (1) single minimizer, (2) multiple minimizers, (3) continuum minimizers, and (4) mixture minimizers.
When reading the table, the row indicates the potential outcome $Y_1$ while the column indicates the potential outcome $Y_0$. The detail of each model can be found in Appendix A.1.1. The subsample size used to compute Table A.2 and Figure A.1 is 120. Figure A.1 shows the evolution of the BN-CI coverage over $k \in [5, 40]$. In all cases, the coverage before the cutoff line $k = \min(40, \frac{0.3b}{m})$ is close to the nominal rate. Figure A.2 shows that the evolution of BN-CI’s coverage against subsample size $b$ is stable.

<table>
<thead>
<tr>
<th>$\tau_n$</th>
<th>0.001, $k = 5$</th>
<th></th>
<th></th>
<th></th>
<th>$\tau_n$</th>
<th>0.002, $k = 10$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.949</td>
<td>0.942</td>
<td>0.948</td>
<td>0.939</td>
<td>(1)</td>
<td>0.971</td>
<td>0.964</td>
<td>0.972</td>
<td>0.952</td>
</tr>
<tr>
<td></td>
<td>(0.176)</td>
<td>(0.167)</td>
<td>(0.152)</td>
<td>(0.169)</td>
<td></td>
<td>(0.186)</td>
<td>(0.174)</td>
<td>(0.160)</td>
<td>(0.187)</td>
</tr>
<tr>
<td>(2)</td>
<td>0.940</td>
<td>0.947</td>
<td>0.947</td>
<td>0.948</td>
<td>(2)</td>
<td>0.967</td>
<td>0.961</td>
<td>0.969</td>
<td>0.972</td>
</tr>
<tr>
<td></td>
<td>(0.155)</td>
<td>(0.140)</td>
<td>(0.116)</td>
<td>(0.166)</td>
<td></td>
<td>(0.162)</td>
<td>(0.147)</td>
<td>(0.126)</td>
<td>(0.184)</td>
</tr>
<tr>
<td>(3)</td>
<td>0.946</td>
<td>0.950</td>
<td>0.955</td>
<td>0.952</td>
<td>(3)</td>
<td>0.967</td>
<td>0.964</td>
<td>0.970</td>
<td>0.964</td>
</tr>
<tr>
<td></td>
<td>(0.135)</td>
<td>(0.122)</td>
<td>(0.061)</td>
<td>(0.106)</td>
<td></td>
<td>(0.138)</td>
<td>(0.127)</td>
<td>(0.069)</td>
<td>(0.118)</td>
</tr>
<tr>
<td>(4)</td>
<td>0.950</td>
<td>0.954</td>
<td>0.947</td>
<td>0.937</td>
<td>(4)</td>
<td>0.970</td>
<td>0.966</td>
<td>0.962</td>
<td>0.961</td>
</tr>
<tr>
<td></td>
<td>(0.185)</td>
<td>(0.177)</td>
<td>(0.171)</td>
<td>(0.165)</td>
<td></td>
<td>(0.205)</td>
<td>(0.200)</td>
<td>(0.191)</td>
<td>(0.186)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\tau_n$</th>
<th>0.004, $k = 20$</th>
<th></th>
<th></th>
<th></th>
<th>$\tau_n$</th>
<th>0.008, $k = 40$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.978</td>
<td>0.971</td>
<td>0.976</td>
<td>0.981</td>
<td>(1)</td>
<td>0.983</td>
<td>0.978</td>
<td>0.965</td>
<td>0.891</td>
</tr>
<tr>
<td></td>
<td>(0.229)</td>
<td>(0.223)</td>
<td>(0.208)</td>
<td>(0.281)</td>
<td></td>
<td>(0.193)</td>
<td>(0.185)</td>
<td>(0.166)</td>
<td>(0.328)</td>
</tr>
<tr>
<td>(2)</td>
<td>0.980</td>
<td>0.974</td>
<td>0.964</td>
<td>0.976</td>
<td>(2)</td>
<td>0.968</td>
<td>0.968</td>
<td>0.963</td>
<td>0.912</td>
</tr>
<tr>
<td></td>
<td>(0.202)</td>
<td>(0.185)</td>
<td>(0.165)</td>
<td>(0.282)</td>
<td></td>
<td>(0.164)</td>
<td>(0.163)</td>
<td>(0.137)</td>
<td>(0.327)</td>
</tr>
<tr>
<td>(3)</td>
<td>0.982</td>
<td>0.975</td>
<td>0.967</td>
<td>0.982</td>
<td>(3)</td>
<td>0.983</td>
<td>0.978</td>
<td>0.966</td>
<td>0.903</td>
</tr>
<tr>
<td></td>
<td>(0.173)</td>
<td>(0.166)</td>
<td>(0.098)</td>
<td>(0.198)</td>
<td></td>
<td>(0.156)</td>
<td>(0.145)</td>
<td>(0.089)</td>
<td>(0.249)</td>
</tr>
<tr>
<td>(4)</td>
<td>0.992</td>
<td>0.987</td>
<td>0.984</td>
<td>0.989</td>
<td>(4)</td>
<td>0.955</td>
<td>0.938</td>
<td>0.948</td>
<td>0.949</td>
</tr>
<tr>
<td></td>
<td>(0.362)</td>
<td>(0.354)</td>
<td>(0.347)</td>
<td>(0.348)</td>
<td></td>
<td>(0.401)</td>
<td>(0.399)</td>
<td>(0.389)</td>
<td>(0.274)</td>
</tr>
</tbody>
</table>
Table A.3: Coverage of 95% \( n \) out of \( n \) bootstrap CI, sample size = 300

<table>
<thead>
<tr>
<th>( \tau_n ) = 0.001, ( k = 5 )</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>( \tau_n ) = 0.002, ( k = 10 )</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.858</td>
<td>0.841</td>
<td>0.847</td>
<td>0.833</td>
<td>(1)</td>
<td>0.872</td>
<td>0.884</td>
<td>0.842</td>
<td>0.864</td>
</tr>
<tr>
<td>(2)</td>
<td>0.868</td>
<td>0.868</td>
<td>0.837</td>
<td>0.820</td>
<td>(2)</td>
<td>0.874</td>
<td>0.877</td>
<td>0.878</td>
<td>0.840</td>
</tr>
<tr>
<td>(3)</td>
<td>0.846</td>
<td>0.814</td>
<td>0.871</td>
<td>0.842</td>
<td>(3)</td>
<td>0.844</td>
<td>0.855</td>
<td>0.879</td>
<td>0.866</td>
</tr>
<tr>
<td>(4)</td>
<td>0.864</td>
<td>0.861</td>
<td>0.841</td>
<td>0.863</td>
<td>(4)</td>
<td>0.884</td>
<td>0.872</td>
<td>0.871</td>
<td>0.886</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \tau_n ) = 0.004, ( k = 20 )</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>( \tau_n ) = 0.008, ( k = 40 )</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.908</td>
<td>0.885</td>
<td>0.867</td>
<td>0.901</td>
<td>(1)</td>
<td>0.929</td>
<td>0.919</td>
<td>0.915</td>
<td>0.927</td>
</tr>
<tr>
<td>(2)</td>
<td>0.901</td>
<td>0.908</td>
<td>0.894</td>
<td>0.881</td>
<td>(2)</td>
<td>0.928</td>
<td>0.924</td>
<td>0.916</td>
<td>0.907</td>
</tr>
<tr>
<td>(3)</td>
<td>0.901</td>
<td>0.881</td>
<td>0.893</td>
<td>0.892</td>
<td>(3)</td>
<td>0.927</td>
<td>0.921</td>
<td>0.909</td>
<td>0.927</td>
</tr>
<tr>
<td>(4)</td>
<td>0.892</td>
<td>0.901</td>
<td>0.892</td>
<td>0.928</td>
<td>(4)</td>
<td>0.917</td>
<td>0.905</td>
<td>0.919</td>
<td>0.938</td>
</tr>
</tbody>
</table>

Each \((i, j)\)-th subplot represents the \((i, j)\)-th model. The dashed line is the coverage of BN-CI with \( b = 120 \) and \( n = 300 \) for quantile index \( \tau \in [1.67\%, 16.67\%] \). The dotted line is the coverage of NN-CI. The horizontal dotted dashed line is the 95% nominal coverage rate, and the vertical dotted dashed line is \( \tau = \min\left( \frac{40}{n}, \frac{0.2b}{n} \right) \).

**Figure A.1:** Coverage across quantiles
Each \((i,j)\)-th subplot represents the \((i,j)\)-th model. The solid line is the coverage for \(b\) out of \(n\) bootstrap CI at \(k = 10\) in which \(b \in [100, 200]\).

**Figure A.2:** Coverage across subsample size

---

Table A.4: Bias of the median-unbiased estimator, sample size = 300. All values are inflated by 100.

<table>
<thead>
<tr>
<th>(\tau_n) = 0.017, (k = 5)</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(\tau_n) = 0.033, (k = 10)</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.372</td>
<td>0.132</td>
<td>0.376</td>
<td>-0.138</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>0.315</td>
<td>0.154</td>
<td>0.049</td>
<td>-0.058</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>-0.109</td>
<td>0.139</td>
<td>0.011</td>
<td>-0.127</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>0.198</td>
<td>0.011</td>
<td>0.100</td>
<td>0.086</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\tau_n) = 0.067, (k = 20)</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(\tau_n) = 0.133, (k = 40)</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>(1)</td>
<td>-0.024</td>
<td>0.024</td>
<td>0.139</td>
<td>-0.075</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>0.079</td>
<td>0.182</td>
<td>0.113</td>
<td>-0.089</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>-0.154</td>
<td>-0.087</td>
<td>0.038</td>
<td>-0.149</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>-0.163</td>
<td>0.030</td>
<td>0.055</td>
<td>-0.033</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table A.5: root-MSE of the median-unbiased estimator, sample size = 300. All values are inflated by 100.

\[
\begin{array}{cccc}
\tau_n = 0.017, k = 5 & (1) & (2) & (3) & (4) \\
(1) & 2.996 & 2.992 & 2.588 & 3.047 \\
(2) & 2.608 & 2.421 & 1.911 & 2.964 \\
(3) & 2.260 & 1.998 & 0.995 & 1.899 \\
(4) & 2.754 & 2.672 & 2.695 & 2.890 \\
\tau_n = 0.033, k = 10 & (1) & (2) & (3) & (4) \\
(1) & 3.194 & 3.218 & 2.868 & 3.462 \\
(2) & 2.748 & 2.770 & 2.351 & 3.226 \\
(3) & 2.397 & 2.141 & 1.222 & 2.212 \\
(4) & 3.253 & 3.291 & 3.301 & 3.598 \\
\frac{\tau_n}{100} = 0.067, k = 20 & (1) & (2) & (3) & (4) \\
(1) & 3.678 & 3.691 & 3.426 & 4.163 \\
(2) & 3.120 & 3.062 & 2.725 & 3.974 \\
(3) & 2.714 & 2.433 & 1.527 & 2.682 \\
(4) & 4.053 & 4.298 & 4.156 & 4.511 \\
\frac{\tau_n}{100} = 0.133, k = 40 & (1) & (2) & (3) & (4) \\
(1) & 4.057 & 4.037 & 4.035 & 5.375 \\
(2) & 3.494 & 3.450 & 3.238 & 5.299 \\
(3) & 3.002 & 2.795 & 2.010 & 3.935 \\
(4) & 5.819 & 5.951 & 6.107 & 8.230 \\
\end{array}
\]

Table A.6: median-bias of the median-unbiased estimator, sample size = 300. All values are inflated by 100.

\[
\begin{array}{cccc}
\tau_n = 0.017, k = 5 & (1) & (2) & (3) & (4) \\
(1) & 0.262 & -0.005 & 0.217 & -0.044 \\
(2) & 0.322 & 0.235 & -0.040 & 0.156 \\
(3) & -0.012 & 0.248 & -0.010 & -0.079 \\
(4) & 0.082 & -0.013 & 0.010 & 0.028 \\
\tau_n = 0.033, k = 10 & (1) & (2) & (3) & (4) \\
(1) & 0.138 & -0.060 & 0.155 & 0.009 \\
(2) & 0.155 & 0.021 & 0.055 & 0.103 \\
(3) & -0.079 & 0.045 & 0.016 & -0.124 \\
(4) & -0.136 & -0.036 & -0.224 & -0.160 \\
\frac{\tau_n}{100} = 0.067, k = 20 & (1) & (2) & (3) & (4) \\
(1) & -0.192 & -0.072 & -0.096 & 0.239 \\
(2) & -0.061 & 0.144 & -0.057 & -0.037 \\
(3) & -0.144 & -0.089 & -0.007 & -0.082 \\
(4) & -0.397 & -0.229 & -0.122 & -0.231 \\
\frac{\tau_n}{100} = 0.133, k = 40 & (1) & (2) & (3) & (4) \\
(1) & -0.164 & -0.271 & 0.098 & 0.228 \\
(2) & -0.092 & -0.076 & 0.066 & 0.271 \\
(3) & -0.041 & -0.044 & -0.010 & 0.010 \\
(4) & -0.553 & -0.510 & -0.196 & 1.542 \\
\end{array}
\]
Table A.7: MAE of the median-unbiased estimator, sample size = 300. All values are inflated by 100.

<table>
<thead>
<tr>
<th>$\tau_n$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>$\tau_n$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.017, k = 5</td>
<td>1.763</td>
<td>1.883</td>
<td>1.512</td>
<td>2.030</td>
<td>0.033, k = 10</td>
<td>2.021</td>
<td>2.218</td>
<td>1.817</td>
<td>2.368</td>
</tr>
<tr>
<td>(1)</td>
<td>1.687</td>
<td>1.633</td>
<td>1.213</td>
<td>1.980</td>
<td>(2)</td>
<td>1.766</td>
<td>1.838</td>
<td>1.520</td>
<td>2.097</td>
</tr>
<tr>
<td>(2)</td>
<td>1.502</td>
<td>1.395</td>
<td>0.617</td>
<td>1.318</td>
<td>(3)</td>
<td>1.594</td>
<td>1.430</td>
<td>0.784</td>
<td>1.487</td>
</tr>
<tr>
<td>(3)</td>
<td>1.765</td>
<td>1.701</td>
<td>1.588</td>
<td>1.861</td>
<td>(4)</td>
<td>2.128</td>
<td>2.167</td>
<td>2.040</td>
<td>2.192</td>
</tr>
<tr>
<td>(4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\tau_n$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>$\tau_n$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.067, k = 20</td>
<td>2.354</td>
<td>2.510</td>
<td>2.320</td>
<td>2.766</td>
<td>0.133, k = 40</td>
<td>2.836</td>
<td>2.754</td>
<td>2.662</td>
<td>3.488</td>
</tr>
<tr>
<td>(1)</td>
<td>2.057</td>
<td>2.116</td>
<td>1.817</td>
<td>2.537</td>
<td>(2)</td>
<td>2.430</td>
<td>2.511</td>
<td>2.195</td>
<td>3.598</td>
</tr>
<tr>
<td>(2)</td>
<td>1.841</td>
<td>1.576</td>
<td>0.994</td>
<td>1.880</td>
<td>(3)</td>
<td>1.868</td>
<td>1.887</td>
<td>1.363</td>
<td>2.549</td>
</tr>
<tr>
<td>(3)</td>
<td>2.776</td>
<td>2.805</td>
<td>2.871</td>
<td>3.117</td>
<td>(4)</td>
<td>3.999</td>
<td>3.936</td>
<td>4.106</td>
<td>5.272</td>
</tr>
<tr>
<td>(4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

To compute the hybrid CI, $\tau_1 := \min(\frac{40}{n} \cdot 0.2b, mn_{0.2b})$ where the spacing parameter $m$ here is 2. To compute the feasible normalizing factor $\hat{\alpha}_n$ for $\tau$, when $k := \tau n \leq 25$, the spacing parameter is 2 and $k'_l = 10$ while $m = 1.2$ and $k'_l = 20$ when $k > 25$. 

The dashed line is the coverage for BN-CI. The dotted line is the coverage for NN-CI. The solid line is the coverage for the hybrid CI. When $b = 1.2$, $n = 300$, and $\tau \in [6.67\%, \, 20\%]$. The horizontal dotted dashed line is the 95% nominal coverage rate. $\tau_1 = 4\%$ and $\tau_2 = 16.75\%$.

**Figure A.3:** Coverage across quantiles

149
For the lower boundary, I use $\tau_n = (0.02, 0.04, 0.06)$ for $n = 300$ to compute the EV index. The subsample size used is the same as in Table A.2.

Table A.8: Coverage of 95% CI, sample size $= 300$.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.946</td>
<td>0.956</td>
<td>0.967</td>
<td>0.972</td>
</tr>
<tr>
<td></td>
<td>(0.605)</td>
<td>(0.551)</td>
<td>(0.431)</td>
<td>(0.497)</td>
</tr>
<tr>
<td>(2)</td>
<td>0.958</td>
<td>0.960</td>
<td>0.964</td>
<td>0.973</td>
</tr>
<tr>
<td></td>
<td>(0.481)</td>
<td>(0.456)</td>
<td>(0.329)</td>
<td>(0.428)</td>
</tr>
<tr>
<td>(3)</td>
<td>0.935</td>
<td>0.940</td>
<td>0.959</td>
<td>0.966</td>
</tr>
<tr>
<td></td>
<td>(0.392)</td>
<td>(0.352)</td>
<td>(0.153)</td>
<td>(0.226)</td>
</tr>
<tr>
<td>(4)</td>
<td>0.950</td>
<td>0.964</td>
<td>0.958</td>
<td>0.953</td>
</tr>
<tr>
<td></td>
<td>(0.570)</td>
<td>(0.514)</td>
<td>(0.438)</td>
<td>(0.303)</td>
</tr>
</tbody>
</table>

The solid line is the coverage for $b$ out of $n$ bootstrap CI at $k = 0$ in which $b \in [100, 200]$.

**Figure A.4**: Coverage across subsample size

Table A.9: Bias of the median-unbiased 0-QTE estimator, sample size $= 300$. All values are inflated by 100.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>-1.639</td>
<td>-3.178</td>
<td>0.201</td>
<td>-0.408</td>
</tr>
<tr>
<td>(2)</td>
<td>-1.635</td>
<td>-0.927</td>
<td>-0.834</td>
<td>-0.145</td>
</tr>
<tr>
<td>(3)</td>
<td>-1.097</td>
<td>-0.436</td>
<td>-0.559</td>
<td>-1.122</td>
</tr>
<tr>
<td>(4)</td>
<td>-3.313</td>
<td>-2.065</td>
<td>-1.116</td>
<td>-1.909</td>
</tr>
</tbody>
</table>

150
Table A.10: root-MSE of the median-unbiased 0-QTE estimator, sample size = 300. All values are inflated by 100.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>19.809</td>
<td>47.957</td>
<td>27.608</td>
<td>10.520</td>
</tr>
<tr>
<td>(2)</td>
<td>16.819</td>
<td>19.857</td>
<td>19.598</td>
<td>9.221</td>
</tr>
<tr>
<td>(3)</td>
<td>10.612</td>
<td>11.724</td>
<td>4.066</td>
<td>4.378</td>
</tr>
<tr>
<td>(4)</td>
<td>16.769</td>
<td>11.378</td>
<td>8.510</td>
<td>6.018</td>
</tr>
</tbody>
</table>

Table A.11: median-bias of the median-unbiased 0-QTE estimator, sample size = 300. All values are inflated by 100.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>-1.623</td>
<td>-0.504</td>
<td>0.966</td>
<td>-1.335</td>
</tr>
<tr>
<td>(2)</td>
<td>-1.881</td>
<td>-1.243</td>
<td>-0.233</td>
<td>-1.130</td>
</tr>
<tr>
<td>(3)</td>
<td>-1.920</td>
<td>-1.365</td>
<td>-0.461</td>
<td>-1.431</td>
</tr>
<tr>
<td>(4)</td>
<td>-2.591</td>
<td>-2.069</td>
<td>-0.925</td>
<td>-1.965</td>
</tr>
</tbody>
</table>

Table A.12: MAE of the median-unbiased 0-QTE estimator, sample size = 300. All values are inflated by 100.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>7.336</td>
<td>7.378</td>
<td>5.705</td>
<td>6.080</td>
</tr>
<tr>
<td>(2)</td>
<td>7.048</td>
<td>7.063</td>
<td>4.015</td>
<td>4.738</td>
</tr>
<tr>
<td>(3)</td>
<td>5.900</td>
<td>4.937</td>
<td>1.997</td>
<td>2.933</td>
</tr>
<tr>
<td>(4)</td>
<td>6.387</td>
<td>5.198</td>
<td>4.241</td>
<td>3.679</td>
</tr>
</tbody>
</table>

Simulation results with $n = 1,000$

Next I consider the QTE estimator with a moderate size sample: 1,000. I am still interested in $k = (5, 10, 20, 40)$ and the corresponding quantile indices become $\tau_n = (0.005, 0.01, 0.015, 0.02)$. The subsample size used in Table A.13 and Figure A.5 is 300. For Figure A.6, the subsample size ranges from 150 to 500.
Table A.13: Coverage of 95% $b$ out of $n$ bootstrap CI, sample size $1,000$

<table>
<thead>
<tr>
<th>$\tau_n = 0.5%$, $k =$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>$\tau_n = 1%$, $k =$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.915</td>
<td>0.918</td>
<td>0.927</td>
<td>0.924</td>
<td>(1)</td>
<td>0.939</td>
<td>0.946</td>
<td>0.957</td>
<td>0.934</td>
</tr>
<tr>
<td></td>
<td>(0.065)</td>
<td>(0.063)</td>
<td>(0.051)</td>
<td>(0.060)</td>
<td></td>
<td>(0.073)</td>
<td>(0.070)</td>
<td>(0.059)</td>
<td>(0.068)</td>
</tr>
<tr>
<td>(2)</td>
<td>0.918</td>
<td>0.930</td>
<td>0.942</td>
<td>0.931</td>
<td>(2)</td>
<td>0.950</td>
<td>0.945</td>
<td>0.955</td>
<td>0.953</td>
</tr>
<tr>
<td></td>
<td>(0.061)</td>
<td>(0.054)</td>
<td>(0.038)</td>
<td>(0.058)</td>
<td></td>
<td>(0.066)</td>
<td>(0.061)</td>
<td>(0.045)</td>
<td>(0.068)</td>
</tr>
<tr>
<td>(3)</td>
<td>0.926</td>
<td>0.933</td>
<td>0.949</td>
<td>0.921</td>
<td>(3)</td>
<td>0.950</td>
<td>0.949</td>
<td>0.954</td>
<td>0.948</td>
</tr>
<tr>
<td></td>
<td>(0.055)</td>
<td>(0.048)</td>
<td>(0.019)</td>
<td>(0.036)</td>
<td></td>
<td>(0.060)</td>
<td>(0.053)</td>
<td>(0.023)</td>
<td>(0.044)</td>
</tr>
<tr>
<td>(4)</td>
<td>0.917</td>
<td>0.902</td>
<td>0.935</td>
<td>0.931</td>
<td>(4)</td>
<td>0.958</td>
<td>0.949</td>
<td>0.957</td>
<td>0.955</td>
</tr>
<tr>
<td></td>
<td>(0.059)</td>
<td>(0.052)</td>
<td>(0.047)</td>
<td>(0.052)</td>
<td></td>
<td>(0.072)</td>
<td>(0.065)</td>
<td>(0.060)</td>
<td>(0.066)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\tau_n = 2%$, $k =$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>$\tau_n = 4%$, $k =$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.966</td>
<td>0.958</td>
<td>0.959</td>
<td>0.957</td>
<td>(1)</td>
<td>0.969</td>
<td>0.968</td>
<td>0.964</td>
<td>0.981</td>
</tr>
<tr>
<td></td>
<td>(0.077)</td>
<td>(0.073)</td>
<td>(0.063)</td>
<td>(0.079)</td>
<td></td>
<td>(0.097)</td>
<td>(0.097)</td>
<td>(0.082)</td>
<td>(0.111)</td>
</tr>
<tr>
<td>(2)</td>
<td>0.961</td>
<td>0.952</td>
<td>0.965</td>
<td>0.958</td>
<td>(2)</td>
<td>0.966</td>
<td>0.966</td>
<td>0.967</td>
<td>0.981</td>
</tr>
<tr>
<td></td>
<td>(0.070)</td>
<td>(0.064)</td>
<td>(0.049)</td>
<td>(0.083)</td>
<td></td>
<td>(0.090)</td>
<td>(0.081)</td>
<td>(0.066)</td>
<td>(0.116)</td>
</tr>
<tr>
<td>(3)</td>
<td>0.966</td>
<td>0.954</td>
<td>0.953</td>
<td>0.954</td>
<td>(3)</td>
<td>0.978</td>
<td>0.971</td>
<td>0.972</td>
<td>0.977</td>
</tr>
<tr>
<td></td>
<td>(0.064)</td>
<td>(0.058)</td>
<td>(0.025)</td>
<td>(0.051)</td>
<td></td>
<td>(0.081)</td>
<td>(0.073)</td>
<td>(0.039)</td>
<td>(0.078)</td>
</tr>
<tr>
<td>(4)</td>
<td>0.957</td>
<td>0.963</td>
<td>0.957</td>
<td>0.966</td>
<td>(4)</td>
<td>0.990</td>
<td>0.993</td>
<td>0.990</td>
<td>0.989</td>
</tr>
<tr>
<td></td>
<td>(0.080)</td>
<td>(0.070)</td>
<td>(0.066)</td>
<td>(0.073)</td>
<td></td>
<td>(0.142)</td>
<td>(0.127)</td>
<td>(0.122)</td>
<td>(0.132)</td>
</tr>
</tbody>
</table>

Table A.14: Coverage of 95% $n$ out of $n$ bootstrap CI, sample size $1,000$

<table>
<thead>
<tr>
<th>$\tau_n = 0.5%$, $k =$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>$\tau_n = 1%$, $k =$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.822</td>
<td>0.832</td>
<td>0.857</td>
<td>0.836</td>
<td>(1)</td>
<td>0.868</td>
<td>0.861</td>
<td>0.859</td>
<td>0.861</td>
</tr>
<tr>
<td></td>
<td>(0.057)</td>
<td>(0.051)</td>
<td>(0.041)</td>
<td>(0.050)</td>
<td></td>
<td>(0.061)</td>
<td>(0.056)</td>
<td>(0.047)</td>
<td>(0.058)</td>
</tr>
<tr>
<td>(2)</td>
<td>0.850</td>
<td>0.841</td>
<td>0.844</td>
<td>0.812</td>
<td>(2)</td>
<td>0.873</td>
<td>0.868</td>
<td>0.867</td>
<td>0.841</td>
</tr>
<tr>
<td></td>
<td>(0.049)</td>
<td>(0.044)</td>
<td>(0.031)</td>
<td>(0.045)</td>
<td></td>
<td>(0.053)</td>
<td>(0.049)</td>
<td>(0.037)</td>
<td>(0.054)</td>
</tr>
<tr>
<td>(3)</td>
<td>0.797</td>
<td>0.814</td>
<td>0.846</td>
<td>0.804</td>
<td>(3)</td>
<td>0.839</td>
<td>0.852</td>
<td>0.876</td>
<td>0.852</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.034)</td>
<td>(0.015)</td>
<td>(0.028)</td>
<td></td>
<td>(0.045)</td>
<td>(0.039)</td>
<td>(0.018)</td>
<td>(0.035)</td>
</tr>
<tr>
<td>(4)</td>
<td>0.827</td>
<td>0.835</td>
<td>0.849</td>
<td>0.858</td>
<td>(4)</td>
<td>0.866</td>
<td>0.866</td>
<td>0.873</td>
<td>0.884</td>
</tr>
<tr>
<td></td>
<td>(0.049)</td>
<td>(0.043)</td>
<td>(0.038)</td>
<td>(0.042)</td>
<td></td>
<td>(0.055)</td>
<td>(0.050)</td>
<td>(0.046)</td>
<td>(0.052)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\tau_n = 2%$, $k =$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>$\tau_n = 4%$, $k =$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.899</td>
<td>0.883</td>
<td>0.879</td>
<td>0.883</td>
<td>(1)</td>
<td>0.914</td>
<td>0.893</td>
<td>0.871</td>
<td>0.896</td>
</tr>
<tr>
<td></td>
<td>(0.066)</td>
<td>(0.063)</td>
<td>(0.055)</td>
<td>(0.066)</td>
<td></td>
<td>(0.074)</td>
<td>(0.072)</td>
<td>(0.065)</td>
<td>(0.077)</td>
</tr>
<tr>
<td>(2)</td>
<td>0.895</td>
<td>0.892</td>
<td>0.885</td>
<td>0.885</td>
<td>(2)</td>
<td>0.910</td>
<td>0.911</td>
<td>0.901</td>
<td>0.894</td>
</tr>
<tr>
<td></td>
<td>(0.057)</td>
<td>(0.054)</td>
<td>(0.043)</td>
<td>(0.062)</td>
<td></td>
<td>(0.063)</td>
<td>(0.062)</td>
<td>(0.051)</td>
<td>(0.072)</td>
</tr>
<tr>
<td>(3)</td>
<td>0.872</td>
<td>0.864</td>
<td>0.896</td>
<td>0.875</td>
<td>(3)</td>
<td>0.912</td>
<td>0.893</td>
<td>0.917</td>
<td>0.903</td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.044)</td>
<td>(0.022)</td>
<td>(0.040)</td>
<td></td>
<td>(0.052)</td>
<td>(0.049)</td>
<td>(0.029)</td>
<td>(0.051)</td>
</tr>
<tr>
<td>(4)</td>
<td>0.894</td>
<td>0.887</td>
<td>0.886</td>
<td>0.902</td>
<td>(4)</td>
<td>0.922</td>
<td>0.922</td>
<td>0.921</td>
<td>0.914</td>
</tr>
<tr>
<td></td>
<td>(0.064)</td>
<td>(0.061)</td>
<td>(0.058)</td>
<td>(0.064)</td>
<td></td>
<td>(0.079)</td>
<td>(0.076)</td>
<td>(0.075)</td>
<td>(0.084)</td>
</tr>
</tbody>
</table>
Each \((i, j)\)-th subplot represents the \((i, j)\)-th model. The dashed line is the coverage of BN-CI with \(b = 300\) and \(n = 1,000\) for quantile index \(\tau \in [0.5\%, 10\%]\). The dotted line is the coverage of NN-CI. The horizontal dotted dashed line is the 95\% nominal coverage rate, and the vertical dotted dashed line is \(\tau = \min \left( \frac{40}{m}, \frac{0.2b}{m} \right)\).

**Figure A.5:** Coverage across quantiles

Each \((i, j)\)-th subplot represents the \((i, j)\)-th model. The solid line is the coverage for \(b\) out of \(n\) bootstrap CI at \(k = 10\) in which \(b \in [150, 500]\).

**Figure A.6:** Coverage across subsample size
Table A.15: Bias of the median-unbiased estimator, sample size = 1,000. All values are inflated by 1,000.

<table>
<thead>
<tr>
<th>$\tau_n$ = 0.005, $k =$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>$\tau_n$ = 0.010, $k =$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>-0.323</td>
<td>0.077</td>
<td>1.079</td>
<td>-0.441</td>
<td>(1)</td>
<td>-0.349</td>
<td>0.021</td>
<td>0.940</td>
<td>-0.107</td>
</tr>
<tr>
<td>0.10</td>
<td>-0.963</td>
<td>0.386</td>
<td>0.206</td>
<td>-1.186</td>
<td>(2)</td>
<td>-0.964</td>
<td>-0.450</td>
<td>0.432</td>
<td>-1.487</td>
</tr>
<tr>
<td>0.20</td>
<td>-0.561</td>
<td>-0.070</td>
<td>-0.154</td>
<td>-0.647</td>
<td>(3)</td>
<td>-0.459</td>
<td>-0.272</td>
<td>0.067</td>
<td>-0.767</td>
</tr>
<tr>
<td>0.40</td>
<td>-0.044</td>
<td>-0.368</td>
<td>0.229</td>
<td>-0.527</td>
<td>(4)</td>
<td>0.230</td>
<td>-0.455</td>
<td>1.002</td>
<td>-0.665</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\tau_n$ = 0.020, $k =$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>$\tau_n$ = 0.040, $k =$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.297</td>
<td>0.674</td>
<td>1.324</td>
<td>-0.213</td>
<td>(1)</td>
<td>0.956</td>
<td>0.123</td>
<td>0.981</td>
<td>-0.933</td>
</tr>
<tr>
<td>0.10</td>
<td>-0.715</td>
<td>-0.185</td>
<td>0.006</td>
<td>-0.652</td>
<td>(2)</td>
<td>0.273</td>
<td>0.450</td>
<td>0.229</td>
<td>-1.097</td>
</tr>
<tr>
<td>0.20</td>
<td>-0.482</td>
<td>-0.261</td>
<td>0.066</td>
<td>-0.328</td>
<td>(3)</td>
<td>-0.523</td>
<td>0.437</td>
<td>-0.010</td>
<td>-0.322</td>
</tr>
<tr>
<td>0.40</td>
<td>0.270</td>
<td>-0.360</td>
<td>0.977</td>
<td>-0.178</td>
<td>(4)</td>
<td>0.328</td>
<td>0.391</td>
<td>1.286</td>
<td>-0.636</td>
</tr>
</tbody>
</table>

Table A.16: root-MSE of the median-unbiased estimator, sample size = 1,000. All values are inflated by 100.

<table>
<thead>
<tr>
<th>$\tau_n$ = 0.005, $k =$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>$\tau_n$ = 0.010, $k =$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>1.348</td>
<td>1.251</td>
<td>1.121</td>
<td>1.273</td>
<td>(1)</td>
<td>1.409</td>
<td>1.338</td>
<td>1.168</td>
<td>1.402</td>
</tr>
<tr>
<td>0.10</td>
<td>1.242</td>
<td>1.083</td>
<td>0.790</td>
<td>1.252</td>
<td>(2)</td>
<td>1.322</td>
<td>1.211</td>
<td>0.924</td>
<td>1.403</td>
</tr>
<tr>
<td>0.20</td>
<td>1.111</td>
<td>0.958</td>
<td>0.384</td>
<td>0.748</td>
<td>(3)</td>
<td>1.130</td>
<td>0.993</td>
<td>0.456</td>
<td>0.889</td>
</tr>
<tr>
<td>0.40</td>
<td>1.194</td>
<td>1.030</td>
<td>0.933</td>
<td>1.072</td>
<td>(4)</td>
<td>1.319</td>
<td>1.209</td>
<td>1.156</td>
<td>1.199</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\tau_n$ = 0.020, $k =$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>$\tau_n$ = 0.040, $k =$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>1.583</td>
<td>1.547</td>
<td>1.363</td>
<td>1.640</td>
<td>(1)</td>
<td>1.803</td>
<td>1.736</td>
<td>1.594</td>
<td>1.941</td>
</tr>
<tr>
<td>0.10</td>
<td>1.389</td>
<td>1.377</td>
<td>1.079</td>
<td>1.611</td>
<td>(2)</td>
<td>1.538</td>
<td>1.483</td>
<td>1.262</td>
<td>1.827</td>
</tr>
<tr>
<td>0.20</td>
<td>1.177</td>
<td>1.103</td>
<td>0.555</td>
<td>1.010</td>
<td>(3)</td>
<td>1.307</td>
<td>1.239</td>
<td>0.666</td>
<td>1.202</td>
</tr>
<tr>
<td>0.40</td>
<td>1.492</td>
<td>1.441</td>
<td>1.438</td>
<td>1.480</td>
<td>(4)</td>
<td>1.866</td>
<td>1.832</td>
<td>1.830</td>
<td>1.935</td>
</tr>
</tbody>
</table>
Table A.17: median-bias of the median-unbiased estimator, sample size = 1,000. All values are inflated by 1,000.

<table>
<thead>
<tr>
<th>( \tau_n )</th>
<th>0.005, ( k = 5 )</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>( \tau_n )</th>
<th>0.010, ( k = 10 )</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>-0.237</td>
<td>-0.347</td>
<td>-0.205</td>
<td>0.759</td>
<td>(1)</td>
<td>-0.483</td>
<td>-0.414</td>
<td>0.430</td>
<td>0.080</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>-0.892</td>
<td>0.703</td>
<td>0.359</td>
<td>-0.511</td>
<td>(2)</td>
<td>-1.326</td>
<td>-0.187</td>
<td>0.559</td>
<td>-0.854</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>-0.101</td>
<td>0.459</td>
<td>-0.077</td>
<td>-0.041</td>
<td>(3)</td>
<td>-0.380</td>
<td>0.156</td>
<td>0.220</td>
<td>-0.447</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>-0.368</td>
<td>-0.188</td>
<td>0.302</td>
<td>-0.309</td>
<td>(4)</td>
<td>0.364</td>
<td>-0.403</td>
<td>0.919</td>
<td>-0.926</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \tau_n )</th>
<th>0.020, ( k = 20 )</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>( \tau_n )</th>
<th>0.040, ( k = 40 )</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>-0.089</td>
<td>0.041</td>
<td>0.639</td>
<td>1.274</td>
<td>(1)</td>
<td>0.158</td>
<td>-0.312</td>
<td>0.725</td>
<td>0.338</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>-1.243</td>
<td>-0.342</td>
<td>-0.315</td>
<td>0.261</td>
<td>(2)</td>
<td>0.617</td>
<td>0.511</td>
<td>-0.131</td>
<td>0.157</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>-0.181</td>
<td>0.585</td>
<td>0.011</td>
<td>0.129</td>
<td>(3)</td>
<td>-0.229</td>
<td>0.316</td>
<td>0.037</td>
<td>-0.391</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>0.119</td>
<td>-0.297</td>
<td>0.592</td>
<td>-0.285</td>
<td>(4)</td>
<td>-0.029</td>
<td>-0.542</td>
<td>0.178</td>
<td>-0.868</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table A.18: MAE of the median-unbiased estimator, sample size = 1,000. All values are inflated by 100.

<table>
<thead>
<tr>
<th>( \tau_n )</th>
<th>0.005, ( k = 5 )</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>( \tau_n )</th>
<th>0.010, ( k = 10 )</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.867</td>
<td>0.798</td>
<td>0.724</td>
<td>0.895</td>
<td>(1)</td>
<td>0.902</td>
<td>0.901</td>
<td>0.765</td>
<td>0.916</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>0.842</td>
<td>0.733</td>
<td>0.525</td>
<td>0.807</td>
<td>(2)</td>
<td>0.887</td>
<td>0.814</td>
<td>0.602</td>
<td>0.893</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>0.729</td>
<td>0.660</td>
<td>0.232</td>
<td>0.509</td>
<td>(3)</td>
<td>0.786</td>
<td>0.688</td>
<td>0.289</td>
<td>0.626</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>0.832</td>
<td>0.686</td>
<td>0.588</td>
<td>0.697</td>
<td>(4)</td>
<td>0.894</td>
<td>0.819</td>
<td>0.789</td>
<td>0.802</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \tau_n )</th>
<th>0.020, ( k = 20 )</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>( \tau_n )</th>
<th>0.040, ( k = 40 )</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>1.050</td>
<td>1.017</td>
<td>0.907</td>
<td>1.140</td>
<td>(1)</td>
<td>1.226</td>
<td>1.180</td>
<td>1.080</td>
<td>1.316</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>0.942</td>
<td>0.966</td>
<td>0.726</td>
<td>1.084</td>
<td>(2)</td>
<td>1.006</td>
<td>0.982</td>
<td>0.794</td>
<td>1.216</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>0.780</td>
<td>0.741</td>
<td>0.373</td>
<td>0.688</td>
<td>(3)</td>
<td>0.895</td>
<td>0.811</td>
<td>0.448</td>
<td>0.840</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>1.020</td>
<td>0.988</td>
<td>0.928</td>
<td>0.968</td>
<td>(4)</td>
<td>1.255</td>
<td>1.241</td>
<td>1.225</td>
<td>1.252</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

To compute the hybrid CI, \( \tau_1 := \min\left(\frac{40}{n}, \frac{0.2k}{mn}\right) \) where the spacing parameter \( m \) here is 2 and \( \tau_2 = \frac{b}{n\sqrt{\log(n)}} \). To compute the feasible normalizing factor \( \hat{\alpha}_n \) for \( \tau \), when \( k := \tau n \leq 50 \), the spacing parameter is 2 and \( k' = 10 \) while \( m = 1.2 \) and \( k' = 20 \) when \( k > 50 \).
The dashed line is the coverage for BN-CI. The dotted line is the coverage for NN-CI. The solid line is the coverage for the hybrid CI. When $b = 300$, $n = 1,000$, and $\tau \in [0.5\%, 15\%]$. The horizontal dotted dashed line is the 95% nominal coverage rate. $\tau_1 = 4\%$ and $\tau_2 = 11.41\%$.

**Figure A.7**: Coverage across quantiles

For the lower boundary: I use $\tau_n = (0.02, 0.04, \cdots, 0.1)$ for $n = 1,000$ to compute the EV index. The subsample size used is the same as in Table A.13.

Table A.19: Coverage of 95% CI, sample size = 1,000.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.963</td>
<td>0.969</td>
<td>0.963</td>
<td>0.963</td>
</tr>
<tr>
<td></td>
<td>(0.200)</td>
<td>(0.172)</td>
<td>(0.119)</td>
<td>(0.145)</td>
</tr>
<tr>
<td>(2)</td>
<td>0.956</td>
<td>0.969</td>
<td>0.973</td>
<td>0.966</td>
</tr>
<tr>
<td></td>
<td>(0.168)</td>
<td>(0.155)</td>
<td>(0.093)</td>
<td>(0.129)</td>
</tr>
<tr>
<td>(3)</td>
<td>0.941</td>
<td>0.963</td>
<td>0.955</td>
<td>0.926</td>
</tr>
<tr>
<td></td>
<td>(0.140)</td>
<td>(0.112)</td>
<td>(0.037)</td>
<td>(0.063)</td>
</tr>
<tr>
<td>(4)</td>
<td>0.920</td>
<td>0.918</td>
<td>0.950</td>
<td>0.938</td>
</tr>
<tr>
<td></td>
<td>(0.139)</td>
<td>(0.116)</td>
<td>(0.099)</td>
<td>(0.085)</td>
</tr>
</tbody>
</table>
The solid line is the coverage for $b$ out of $n$ bootstrap CI at $k = 0$ in which $b \in [150, 500]$.

**Figure A.8:** Coverage across subsample size

Table A.20: Bias of the median-unbiased 0-QTE estimator, sample size = 1,000. All values are inflated by 100.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>-0.475</td>
<td>-0.693</td>
<td>0.092</td>
<td>-0.023</td>
</tr>
<tr>
<td>(2)</td>
<td>-0.586</td>
<td>-0.301</td>
<td>-0.046</td>
<td>-0.287</td>
</tr>
<tr>
<td>(3)</td>
<td>-0.813</td>
<td>-0.754</td>
<td>-0.294</td>
<td>-0.675</td>
</tr>
<tr>
<td>(4)</td>
<td>-0.831</td>
<td>-0.924</td>
<td>-0.637</td>
<td>-0.870</td>
</tr>
</tbody>
</table>

Table A.21: root-MSE of the median-unbiased 0-QTE estimator, sample size = 1,000. All values are inflated by 100.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>6.162</td>
<td>5.751</td>
<td>3.711</td>
<td>4.096</td>
</tr>
<tr>
<td>(2)</td>
<td>5.502</td>
<td>4.651</td>
<td>2.809</td>
<td>3.450</td>
</tr>
<tr>
<td>(3)</td>
<td>4.042</td>
<td>3.140</td>
<td>1.152</td>
<td>1.859</td>
</tr>
<tr>
<td>(4)</td>
<td>4.473</td>
<td>3.836</td>
<td>3.049</td>
<td>2.862</td>
</tr>
</tbody>
</table>
Table A.22: median-bias of the median-unbiased 0-QTE estimator, sample size = 1,000. All values are inflated by 100.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>-0.634</td>
<td>-0.861</td>
<td>0.348</td>
<td>-0.322</td>
</tr>
<tr>
<td>(2)</td>
<td>-0.623</td>
<td>-0.304</td>
<td>-0.094</td>
<td>-0.670</td>
</tr>
<tr>
<td>(3)</td>
<td>-1.258</td>
<td>-0.946</td>
<td>-0.305</td>
<td>-0.705</td>
</tr>
<tr>
<td>(4)</td>
<td>-1.135</td>
<td>-0.931</td>
<td>-0.728</td>
<td>-0.937</td>
</tr>
</tbody>
</table>

Table A.23: MAE of the median-unbiased 0-QTE estimator, sample size = 1,000. All values are inflated by 100.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>4.046</td>
<td>3.736</td>
<td>2.413</td>
<td>2.616</td>
</tr>
<tr>
<td>(2)</td>
<td>3.546</td>
<td>3.029</td>
<td>1.785</td>
<td>2.277</td>
</tr>
<tr>
<td>(3)</td>
<td>2.889</td>
<td>2.189</td>
<td>0.711</td>
<td>1.257</td>
</tr>
<tr>
<td>(4)</td>
<td>3.048</td>
<td>2.429</td>
<td>1.910</td>
<td>1.918</td>
</tr>
</tbody>
</table>

Simulation results with $n = 5,000$

Table A.24: Bias of the median-unbiased estimator, sample size = 5,000. All values are inflated by 1,000.

<table>
<thead>
<tr>
<th>$\tau_n = 0.001$, $k =$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>$\tau_n = 0.002$, $k =$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>-0.043</td>
<td>0.179</td>
<td>0.021</td>
<td>-0.483</td>
<td>(1)</td>
<td>0.447</td>
<td>0.317</td>
<td>-0.030</td>
<td>-0.843</td>
</tr>
<tr>
<td>(2)</td>
<td>-0.247</td>
<td>-0.833</td>
<td>0.129</td>
<td>-0.589</td>
<td>(2)</td>
<td>-0.407</td>
<td>-0.798</td>
<td>0.077</td>
<td>-0.916</td>
</tr>
<tr>
<td>(3)</td>
<td>-0.194</td>
<td>-0.123</td>
<td>-0.087</td>
<td>-0.421</td>
<td>(3)</td>
<td>-0.398</td>
<td>-0.322</td>
<td>-0.056</td>
<td>-0.235</td>
</tr>
<tr>
<td>(4)</td>
<td>-0.662</td>
<td>-0.962</td>
<td>-0.106</td>
<td>-0.386</td>
<td>(4)</td>
<td>-0.691</td>
<td>-0.608</td>
<td>-0.034</td>
<td>-0.411</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\tau_n = 0.004$, $k =$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>$\tau_n = 0.008$, $k =$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.156</td>
<td>0.498</td>
<td>0.098</td>
<td>-0.586</td>
<td>(1)</td>
<td>0.392</td>
<td>0.358</td>
<td>-0.265</td>
<td>-0.245</td>
</tr>
<tr>
<td>(2)</td>
<td>-0.172</td>
<td>-0.727</td>
<td>0.123</td>
<td>-0.757</td>
<td>(2)</td>
<td>-0.209</td>
<td>-0.629</td>
<td>-0.009</td>
<td>-0.090</td>
</tr>
<tr>
<td>(3)</td>
<td>-0.390</td>
<td>-0.026</td>
<td>-0.063</td>
<td>-0.417</td>
<td>(3)</td>
<td>-0.278</td>
<td>-0.056</td>
<td>-0.144</td>
<td>-0.257</td>
</tr>
<tr>
<td>(4)</td>
<td>-0.877</td>
<td>-0.180</td>
<td>0.326</td>
<td>-0.501</td>
<td>(4)</td>
<td>-0.280</td>
<td>-0.361</td>
<td>0.150</td>
<td>-0.192</td>
</tr>
</tbody>
</table>
Table A.25: root-MSE of the median-unbiased estimator, sample size = 5,000. All values are inflated by 1,000.

<table>
<thead>
<tr>
<th>$\tau_n$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>$\tau_n$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001, k = 5</td>
<td>5.806</td>
<td>4.715</td>
<td>3.553</td>
<td>4.202</td>
<td>0.002, k = 10</td>
<td>5.924</td>
<td>5.190</td>
<td>3.816</td>
<td>4.733</td>
</tr>
<tr>
<td></td>
<td>5.259</td>
<td>4.284</td>
<td>2.571</td>
<td>4.057</td>
<td></td>
<td>5.447</td>
<td>4.410</td>
<td>2.858</td>
<td>4.608</td>
</tr>
<tr>
<td></td>
<td>5.112</td>
<td>3.530</td>
<td>1.151</td>
<td>2.489</td>
<td></td>
<td>5.211</td>
<td>3.897</td>
<td>1.291</td>
<td>2.869</td>
</tr>
<tr>
<td></td>
<td>4.574</td>
<td>3.676</td>
<td>2.830</td>
<td>3.113</td>
<td></td>
<td>4.908</td>
<td>3.828</td>
<td>3.296</td>
<td>3.707</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\tau_n$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>$\tau_n$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.004, k = 20</td>
<td>6.310</td>
<td>5.429</td>
<td>4.406</td>
<td>5.364</td>
<td>0.008, k = 40</td>
<td>6.820</td>
<td>5.904</td>
<td>4.971</td>
<td>6.109</td>
</tr>
<tr>
<td></td>
<td>5.659</td>
<td>4.675</td>
<td>3.310</td>
<td>5.033</td>
<td></td>
<td>5.804</td>
<td>5.231</td>
<td>3.802</td>
<td>5.674</td>
</tr>
<tr>
<td></td>
<td>5.052</td>
<td>3.962</td>
<td>1.565</td>
<td>3.130</td>
<td></td>
<td>5.070</td>
<td>4.110</td>
<td>1.848</td>
<td>3.638</td>
</tr>
<tr>
<td></td>
<td>5.344</td>
<td>4.289</td>
<td>3.890</td>
<td>4.381</td>
<td></td>
<td>5.399</td>
<td>5.050</td>
<td>4.567</td>
<td>5.029</td>
</tr>
</tbody>
</table>

Table A.26: median-bias of the median-unbiased estimator, sample size = 5,000. All values are inflated by 1,000.

<table>
<thead>
<tr>
<th>$\tau_n$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>$\tau_n$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001, k = 5</td>
<td>-0.247</td>
<td>-0.013</td>
<td>-0.337</td>
<td>-0.190</td>
<td>0.002, k = 10</td>
<td>0.247</td>
<td>0.243</td>
<td>-0.332</td>
<td>-0.745</td>
</tr>
<tr>
<td></td>
<td>-0.039</td>
<td>-0.684</td>
<td>0.023</td>
<td>-0.261</td>
<td></td>
<td>-0.354</td>
<td>-0.724</td>
<td>-0.030</td>
<td>-0.745</td>
</tr>
<tr>
<td></td>
<td>0.110</td>
<td>0.010</td>
<td>-0.017</td>
<td>-0.246</td>
<td></td>
<td>-0.192</td>
<td>-0.203</td>
<td>-0.019</td>
<td>-0.138</td>
</tr>
<tr>
<td></td>
<td>-0.622</td>
<td>-0.758</td>
<td>0.066</td>
<td>-0.174</td>
<td></td>
<td>-0.575</td>
<td>-0.303</td>
<td>0.158</td>
<td>-0.111</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\tau_n$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>$\tau_n$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.004, k = 20</td>
<td>0.186</td>
<td>0.244</td>
<td>-0.175</td>
<td>-0.352</td>
<td>0.008, k = 40</td>
<td>0.342</td>
<td>0.375</td>
<td>-0.503</td>
<td>-0.291</td>
</tr>
<tr>
<td></td>
<td>0.132</td>
<td>-0.610</td>
<td>0.039</td>
<td>-0.411</td>
<td></td>
<td>-0.228</td>
<td>-0.661</td>
<td>0.152</td>
<td>0.105</td>
</tr>
<tr>
<td></td>
<td>0.073</td>
<td>-0.031</td>
<td>-0.008</td>
<td>-0.404</td>
<td></td>
<td>-0.315</td>
<td>-0.069</td>
<td>-0.154</td>
<td>-0.208</td>
</tr>
<tr>
<td></td>
<td>-1.020</td>
<td>-0.002</td>
<td>0.315</td>
<td>-0.296</td>
<td></td>
<td>-0.179</td>
<td>-0.391</td>
<td>0.085</td>
<td>-0.080</td>
</tr>
</tbody>
</table>
Table A.27: MAE of the median-unbiased estimator, sample size = 5,000. All values are inflated by 1,000.

<table>
<thead>
<tr>
<th>( \tau_n )</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>( \tau_n )</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001, ( k = 5 )</td>
<td>3.916</td>
<td>3.073</td>
<td>2.209</td>
<td>2.783</td>
<td>0.002, ( k = 10 )</td>
<td>4.091</td>
<td>3.678</td>
<td>2.453</td>
<td>2.979</td>
</tr>
<tr>
<td>(1)</td>
<td>3.658</td>
<td>2.881</td>
<td>1.644</td>
<td>2.715</td>
<td>(2)</td>
<td>3.542</td>
<td>2.976</td>
<td>1.994</td>
<td>3.099</td>
</tr>
<tr>
<td>(2)</td>
<td>3.440</td>
<td>2.349</td>
<td>0.743</td>
<td>1.649</td>
<td>(3)</td>
<td>3.507</td>
<td>2.654</td>
<td>0.863</td>
<td>1.875</td>
</tr>
<tr>
<td>(3)</td>
<td>3.098</td>
<td>2.380</td>
<td>1.889</td>
<td>2.169</td>
<td>(4)</td>
<td>3.397</td>
<td>2.492</td>
<td>2.131</td>
<td>2.506</td>
</tr>
<tr>
<td>(4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \tau_n )</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>( \tau_n )</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>0.004, ( k = 20 )</td>
<td>4.407</td>
<td>3.901</td>
<td>2.925</td>
<td>3.599</td>
<td>0.008, ( k = 40 )</td>
<td>4.634</td>
<td>4.084</td>
<td>3.393</td>
<td>4.172</td>
</tr>
<tr>
<td>(1)</td>
<td>3.679</td>
<td>3.158</td>
<td>2.242</td>
<td>3.247</td>
<td>(2)</td>
<td>3.806</td>
<td>3.814</td>
<td>2.534</td>
<td>3.813</td>
</tr>
<tr>
<td>(2)</td>
<td>3.635</td>
<td>2.673</td>
<td>1.027</td>
<td>2.169</td>
<td>(3)</td>
<td>3.636</td>
<td>2.798</td>
<td>1.219</td>
<td>2.425</td>
</tr>
<tr>
<td>(3)</td>
<td>3.837</td>
<td>2.860</td>
<td>2.608</td>
<td>3.035</td>
<td>(4)</td>
<td>3.724</td>
<td>3.478</td>
<td>3.119</td>
<td>3.437</td>
</tr>
<tr>
<td>(4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

To compute the hybrid CI, \( \tau_1 := \min\left(\frac{40}{n}, \frac{0.28}{mn}\right) \) where the spacing parameter \( m \) here is 2 and \( \tau_2 = \frac{b}{n\sqrt{\log(n)}} \). To compute the feasible normalizing factor \( \hat{\alpha}_n \) for \( \tau \), when \( k := \tau n \leq 100 \), the spacing parameter is 2 and \( k' = 10 \) while \( m = 1.2 \) and \( k' = 20 \) when \( k > 100 \).
The dashed line is the coverage for BN-CI. The dotted line is the coverage for NN-CI. The solid line is the coverage for the hybrid CI. When $b = 1,000$, $n = 5,000$, and $\tau \in [0.1\%, 8\%]$. The horizontal dotted dashed line is the 95% nominal coverage rate. $\tau_1 = 0.8\%$ and $\tau_2 = 6.85\%$.

**Figure A.9**: Coverage across quantiles

Next are the finite sample performance of the median-unbiased point estimator.

Table A.28: Bias of the median-unbiased 0-QTE estimator, sample size = 5,000. All values are inflated by 100.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>-0.148</td>
<td>-0.282</td>
<td>0.015</td>
<td>0.105</td>
</tr>
<tr>
<td>(2)</td>
<td>0.006</td>
<td>-0.120</td>
<td>-0.021</td>
<td>0.063</td>
</tr>
<tr>
<td>(3)</td>
<td>0.188</td>
<td>-0.082</td>
<td>-0.056</td>
<td>0.027</td>
</tr>
<tr>
<td>(4)</td>
<td>-0.100</td>
<td>-0.284</td>
<td>-0.058</td>
<td>-0.086</td>
</tr>
</tbody>
</table>

Table A.29: root-MSE of the median-unbiased 0-QTE estimator, sample size = 5,000. All values are inflated by 100.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>4.388</td>
<td>2.944</td>
<td>1.861</td>
<td>2.549</td>
</tr>
<tr>
<td>(2)</td>
<td>3.416</td>
<td>2.453</td>
<td>1.360</td>
<td>2.149</td>
</tr>
<tr>
<td>(3)</td>
<td>2.457</td>
<td>1.646</td>
<td>0.470</td>
<td>0.959</td>
</tr>
<tr>
<td>(4)</td>
<td>2.512</td>
<td>1.687</td>
<td>1.205</td>
<td>1.282</td>
</tr>
</tbody>
</table>
Table A.30: median-bias of the median-unbiased 0-QTE estimator, sample size = 5,000. All values are inflated by 100.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>-0.246</td>
<td>-0.251</td>
<td>0.093</td>
<td>0.041</td>
</tr>
<tr>
<td>(2)</td>
<td>-0.134</td>
<td>-0.136</td>
<td>-0.005</td>
<td>-0.099</td>
</tr>
<tr>
<td>(3)</td>
<td>-0.189</td>
<td>-0.291</td>
<td>-0.082</td>
<td>-0.059</td>
</tr>
<tr>
<td>(4)</td>
<td>-0.382</td>
<td>-0.374</td>
<td>-0.109</td>
<td>-0.177</td>
</tr>
</tbody>
</table>

Table A.31: MAE of the median-unbiased 0-QTE estimator, sample size = 5,000. All values are inflated by 100.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>2.287</td>
<td>1.687</td>
<td>1.076</td>
<td>1.498</td>
</tr>
<tr>
<td>(2)</td>
<td>1.998</td>
<td>1.478</td>
<td>0.849</td>
<td>1.404</td>
</tr>
<tr>
<td>(3)</td>
<td>1.548</td>
<td>1.062</td>
<td>0.300</td>
<td>0.604</td>
</tr>
<tr>
<td>(4)</td>
<td>1.637</td>
<td>1.024</td>
<td>0.750</td>
<td>0.830</td>
</tr>
</tbody>
</table>
A.1.4 Proof of Theorem 2.3.1

Before starting the proof, I first state a maximal inequality which is derived in Chernozhukov et al. (2014). See Corollary 5.1 in their paper. \((X_1, \ldots, X_n)\) is a sequence of i.i.d random variables taking values in a measurable space \((S, \mathcal{S})\) with common distribution \(P\). \(\mathcal{F}\) is a generic class of measurable function \(S \rightarrow \mathbb{R}\) with an envelope function \(F\). Let \(\sigma^2 > 0\) be any positive constant such that

\[
\sup_{f \in \mathcal{F}} Pf^2 \leq \sigma^2 \leq \|F\|_{L_2}^2 \quad \text{and} \quad M = \max_{1 \leq i \leq n} F(X_i).
\]

Lemma A.1.1. If \(F \in L^2(P)\) and suppose that there exist constants \(a \geq e\) and \(v \geq 1\) such that the following uniform entropy condition holds:

\[
\sup_{Q} N(\varepsilon \|F\|_{Q,2}, \mathcal{F}, \cdot \|_{Q,2}) \leq \left(\frac{a}{\varepsilon}\right)^v, \quad \forall \varepsilon \in (0, 1],
\]

then

\[
\mathbb{E}\|\sqrt{n}(P_n - P)\|_F \leq \sqrt{v\sigma^2 \log \left(\frac{a\|F\|_{P,2}}{\sigma}\right)} + \frac{v\|M\|_2}{\sqrt{n}} \log \left(\frac{a\|F\|_{P,2}}{\sigma}\right).
\]

Throughout the appendix, for simplicity of notation, I call a term \(U_n(k) = o_p^\ast(r_n)\) \((O_p^\ast(r_n))\) if

\[
\sup_{k \in [\kappa_1, \kappa_2]} \frac{|U_n(k)|}{r_n} = o_p(1)(O_p(1))
\]

for some fixed positive constants \(\kappa_1\) and \(\kappa_2\).

Now I return to the proof of Theorem 2.3.1. Let \(\hat{\Delta}_{1,n}(k) = \lambda_{1,n}(k)(\hat{q}_1(k\tau_n) - q_1(k\tau_n))\) be the maximizer of the rescaled objective function, that is,

\[
\hat{\Delta}_{1,n}(k) = \arg \min_{\Delta \in \mathbb{R}} -\hat{W}_n(k)\Delta(k) + \hat{G}_n(\Delta, k)
\]

(A.1.1)
where
\[
\hat{W}_n(k) = \frac{1}{\sqrt{nk^2}} \sum_{i=1}^{n} \frac{D_i}{P(X_i)} (k \tau_n - 1\{Y_i \leq q_1(k \tau_n)\}),
\]
\[
\hat{G}_n(\Delta, k) = \frac{1}{\sqrt{nk^2}} \sum_{i=1}^{n} \frac{D_i}{P(X_i)} \int_{0}^{\Delta} \left(1 \left\{Y_i \leq q_1(k \tau_n) + \frac{s}{\lambda_{1,n}(k)} \right\} - 1 \left\{Y_i \leq q_1(k \tau_n) \right\} \right) ds.
\]

The proof of the first part of the theorem is divided into three steps. In the first step, by defining
\[
R_n(\Delta, k) = \hat{G}_n(\Delta, k) - \frac{\Delta^2}{2},
\]
I show that
\[
\sup_{|\Delta| \leq M, k \in [\kappa_1, \kappa_2]} |R_n(\Delta, k)| = o_p(1). \tag{A.1.2}
\]
In the second step, I show that
\[
\hat{W}_n(k) = W_n(k) + o^*_p(1)
\]
where
\[
W_n(k) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{i,1,n}(k)
\]
and
\[
\phi_{i,1,n}(k) = \frac{1}{\sqrt{k^2 \tau_n}} \left[ \frac{D_i}{P(X_i)} T_{i,1,n}(k) - \frac{E(T_{i,1,n}(k) | X_i)}{P(X_i)} (D_i - P(X_i)) \right].
\]
In the third step, I show that \{\hat{W}_n(k) : k \in [\kappa_1, \kappa_2]\} is tight. This implies that \{\hat{W}_n(k) : k \in [\kappa_1, \kappa_2]\} is tight too. Given the tightness of \{\hat{W}_n(k) : k \in [\kappa_1, \kappa_2]\} and (A.1.2), I can apply a generalized version of the Convexity lemma in Pollard (1991a) proved in Lemma 2 of Chernozhukov (2000), I can conclude that
\[
\hat{\Delta}_{n,1}(k) = \hat{W}_n(k) + o^*_p(1) = W_n(k) + o^*_p(1)
\]
and \( \{ \hat{\Delta}_{n,1}(k) : k \in [\kappa_1, \kappa_2] \} \) is tight. Similarly, I can show that
\[
\lambda_{0,n}(k)(\hat{q}_0(k\tau_n) - q_0(k\tau_n)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{0,i,n}(k) + o_p^*(1)
\]
where
\[
\phi_{0,i,n}(k) = \frac{1}{\sqrt{k\tau_n}} \left[ \frac{1 - D_i}{1 - P(X_i)} T_{i,0,n}(k) + \frac{E(T_{i,0,n}(k)|X_i)}{1 - P(X_i)} (D_i - P(X_i)) \right]
\]
and that the stochastic process \( \{ \phi_{0,i,n}(k) : k \in [\kappa_1, \kappa_2] \} \) is tight. This concludes the first half of the results in Theorem 2.3.1.

Step 1.
Define
\[
G_n(\Delta, k) = \frac{1}{\sqrt{n\kappa\tau_n}} \sum_{i=1}^{n} \frac{D_i}{P(X_i)} \int_{0}^{\Delta} \left( 1 \left\{ Y_i \leq q_1(k\tau_n) + \frac{M}{\lambda_{1,n}(k)} \right\} - 1 \left\{ Y_i \leq q_1(k\tau_n) \right\} \right) ds.
\]

By Lemma 1 in Hirano et al. (2003), \( \sup_x |\hat{P}(x) - P(x)| = o_p(1) \). Thus uniformly over \( |\Delta| \leq M \),

\[
|\hat{G}_n(\Delta, k) - G_n(\Delta, k)| \leq o_p(1) \left[ \frac{M}{\sqrt{n\kappa\tau_n}} \sum_{i=1}^{n} \left( 1 \left\{ Y_i \leq q_1(k\tau_n) + \frac{M}{\lambda_{1,n}(k)} \right\} - 1 \left\{ Y_i \leq q_1(k\tau_n) \right\} \right) \right.
\]
\[
+ \left. \left( 1 \left\{ Y_i \leq q_1(k\tau_n) \right\} - 1 \left\{ Y_i \leq q_1(k\tau_n) - \frac{M}{\lambda_{1,n}(k)} \right\} \right) \right]
\]
\[
\leq o_p(1) \| \sqrt{n} P_n f \|_{\mathcal{F}_{1,n}} \leq o_p(1) (\| \sqrt{n} (P_n - P) \|_{\mathcal{F}_{1,n}} + \sqrt{n} \| P f \|_{\mathcal{F}_{1,n}})
\]

where
\[
\mathcal{F}_{1,n} = \left\{ \frac{1}{\sqrt{\tau_n}} \left( 1 \left\{ Y_i \leq q_1(k\tau_n) + \frac{M}{\lambda_{1,n}(k)} \right\} - 1 \left\{ Y_i \leq q_1(k\tau_n) \right\} \right) \right.
\]
\[
+ \left. \left( 1 \left\{ Y_i \leq q_1(k\tau_n) \right\} - 1 \left\{ Y_i \leq q_1(k\tau_n) - \frac{M}{\lambda_{1,n}(k)} \right\} \right) , k \in [\kappa_1, \kappa_2] \right\}.
\]

165
with an envelope function $F_{1,n}$ such that

$$F_{1,n} = \frac{1}{\sqrt{\tau_n}} \left( \mathbb{1} \left\{ Y_i \leq q_1(\kappa_2 \tau_n) + \frac{M}{\lambda_{1,n}} \right\} - \mathbb{1} \left\{ Y_i \leq q_1(\kappa_1 \tau_n) \right\} \right) + \mathbb{1} \left\{ Y_i \leq q_1(\kappa_2 \tau_n) \right\} - \mathbb{1} \left\{ Y_i \leq q_1(\kappa_1 \tau_n) - \frac{M}{\lambda_{1,n}} \right\}.$$

I observe that $||F_{1,n}||_{P,2} \leq C < \infty$ and $M_{1,n} = \max_{1 \leq i \leq n} F_{1,n}(U_i) \leq \frac{2}{\sqrt{\tau_n}}$ where $U_i = \{Y_i, D_i, X_i\}$. To see the second observation, I note that $f(q_1(k \tau_n))$ is monotone in $k$ for large $n$ and $k \in [\kappa_1, \kappa_2]$. Hence $\lambda_{1,n}(k) \geq \tilde{\lambda}_{1,n} := \frac{\sqrt{\pi}}{\sqrt{\tau_{2,n}}} f_1(q_1(k \tau_n))$ where $\tilde{k} = \kappa_1$ or $\kappa_2$ depends on whether $f_1$ is monotone decreasing or increasing at the tail.

Furthermore, $q_1(\tau_n)$ and $\lambda_{1,n}(\cdot)$ are monotone. So by repeatedly using Lemma 2.6.18 (iv), (v), and (viii) of Van der Vaart and Wellner (1996), I have

$$\sup_Q N(\varepsilon ||F_{1,n}||_{Q,2}, \mathcal{F}_{1,n}, ||\cdot||_{Q,2}) \leq \left( \frac{a}{\varepsilon} \right)^v, \quad \forall \varepsilon \in (0, 1].$$

By Lemma A.1.1 with $\sigma = ||F_{1,n}||_{P,2}$, I have

$$\mathbb{E}||\sqrt{n}(\mathcal{P}_n - P)||_{F_{1,n}} \lesssim ||F_{1,n}||_{P,2} + \frac{1}{\sqrt{\tau_n n}} = O(1)$$

and thus

$$||\sqrt{n}(\mathcal{P}_n - P)||_{F_{1,n}} = O_p(1). \quad (A.1.4)$$

I next want to show $\sqrt{n} |Pf|_{F_{1,n}} = O(1)$. In fact, I have

$$\sqrt{n} |Pf|_{F_{1,n}} \lesssim \sup_{k \in [\kappa_1, \kappa_2]} \left( \frac{f_1(q_1(k \tau_n) + \frac{\tilde{M}}{\lambda_{1,n}(k)})}{f_1(q_1(k \tau_n))} + \frac{f_1(q_1(k \tau_n) - \frac{\tilde{M}}{\lambda_{1,n}(k)})}{f_1(q_1(k \tau_n))} \right)$$

where $\tilde{M}$ is between zero and $M$. Since $\tau_n n \to \infty$, for any constant $l > 1$ independent of $k$, there exists $N_0 > 0$ independent of $k$ such that for $n > N_0$,

$$\frac{\tilde{M}}{\lambda_{1,n}(k)} = \frac{\tilde{M}(q_1(lk \tau_n) - q_1(k \tau_n))}{\sqrt{n \tau_n} \int_k^l \frac{f(q_1(k \tau_n))}{f(q_1(\tau_n))} dt} \leq (q_1(l k \tau_n) - q_1(k \tau_n)). \quad (A.1.5)$$
Therefore, if \( f_1 \) is monotone increasing at its tail,
\[
\sup_{k \in [\kappa_1, \kappa_2]} \frac{f_1 \left( q_1(k\tau_n) + \frac{s}{\lambda_{1,n}(k)} \right)}{f_1(q_1(k\tau_n))} + \frac{f_1 \left( q_1(k\tau_n) - \frac{s}{\lambda_{1,n}(k)} \right)}{f_1(q_1(k\tau_n))} \leq \frac{f_1(q_1(l\kappa_2\tau_n))}{f_1(q_1(k_1\tau_n))} + 1 = O(1).
\]

Similar argument shows \( \sup_{k \in [\kappa_1, \kappa_2]} \frac{f_1 \left( q_1(k\tau_n) + \frac{s}{\lambda_{1,n}(k)} \right)}{f_1(q_1(k\tau_n))} + \frac{f_1 \left( q_1(k\tau_n) - \frac{s}{\lambda_{1,n}(k)} \right)}{f_1(q_1(k\tau_n))} = O(1) \) when \( f_1 \) is monotone decreasing at its tail. So I obtain the desired result that
\[
\sqrt{n} \left| \mathbb{E} f \right|_{\mathcal{F}_{n,n}} = O_p(1). \tag{A.1.6}
\]

Combining (A.1.1), (A.1.4), and (A.1.6), I have
\[
\sup_{\Delta, k} \left| \hat{G}_n(\Delta, k) - G_n(\Delta, k) \right| = o_p(1). \tag{A.1.7}
\]

Next, I want to show \( G_n(\Delta, k) \rightarrow \frac{\Delta^2}{2} \) uniformly in \( |\Delta| \leq M \) and \( k \in [\kappa_1, \kappa_2] \). It suffices to show
\[
\sup_{|\Delta| \leq M, k \in [\kappa_1, \kappa_2]} \left| \mathbb{E} G_n(\Delta, k) - \frac{\Delta^2}{2} \right| = o(1) \tag{A.1.8}
\]
and
\[
\sup_{|\Delta| \leq M, k \in [\kappa_1, \kappa_2]} \left| G_n(\Delta, k) - \mathbb{E} G_n(\Delta, k) \right| = o_p(1). \tag{A.1.9}
\]

For (A.1.8), I have
\[
\mathbb{E} G_n(\Delta, k) = \frac{n}{\sqrt{nk\tau_n}} \int_0^\Delta \left( F_1 \left( q_1(k\tau_n) + \frac{s}{\lambda_{1,n}(k)} \right) - F_1(q_1(k\tau_n)) \right) \, ds
\]
\[
= \frac{\Delta^2}{2} \frac{f_1 \left( q_1(k\tau_n) + \frac{s(\kappa_1, \kappa_2)}{\lambda_{1,n}(k)} \right)}{f_1(q_1(k\tau_n))}.
\]

By (C.0.27), for any \( l > 1 \), there exists \( N_0 > 1 \) independent of \( k \) such that for \( n > N_0 \), if \( f_1 \) is monotone increasing at its lower tail,
\[
\frac{f_1 \left( q_1(k\tau_n) + \frac{s(\kappa_1, \kappa_2)}{\lambda_{1,n}(k)} \right)}{f_1(q_1(k\tau_n))} \in \left( \frac{f_1(q_1(k\tau_n))}{f_1(q_1(k\tau_n))}, \frac{f_1(q_1(lk\tau_n))}{f_1(q_1(k\tau_n))} \right).
\]
and if $f_1$ is monotone decreasing in its lower tail,

\[
\frac{f_1\left(q_1(k\tau_n) + \frac{\hat{\Delta}(k, \tau_n)}{\lambda_1,n(k)}\right)}{f_1(q_1(k\tau_n))} \in \left(\frac{f_1(q_1(lk\tau_n))}{f_1(q_1(k\tau_n))}, \frac{f_1\left(q_1(\frac{k}{7}\tau_n)\right)}{f_1(q_1(k\tau_n))}\right).
\]

By first letting $n \to \infty$ and then $l \to 1$, both sides converge to 1 uniformly over $k \in [\kappa_1, \kappa_2]$. This implies $\frac{f_1\left(q_1(k\tau_n) + \frac{\hat{\Delta}(k, \tau_n)}{\lambda_1,n(k)}\right)}{f_1(q_1(k\tau_n))} \to 1$ uniformly in $k$. Therefore, $E G_n(\Delta, k) \to \frac{\Delta^2}{2}$ uniformly in $\Delta$ and $k$.

For (A.1.9), I have

\[
G_n(\Delta, k) - E G_n(\Delta, k) = \sqrt{n}(P_n - P)f \text{ for } f \in F_{2,n}
\]

\[
F_{2,n} = \left\{ \frac{1}{\sqrt{\tau_n}} \frac{D_i}{P(X_i)} \int_0^{\Delta} \left(1 \left\{ Y_i \leq q_1(k\tau_n) + \frac{s}{\lambda_1,n(k)} \right\} - 1 \left\{ Y_i \leq q_1(k\tau_n) \right\} \right) ds, |\Delta| < M, k \in [\kappa_1, \kappa_2] \right\}
\]

where

with an envelope function $F_{2,n} = \frac{D_i}{P(X_i)} F_{1,n}$. I note that $\|F_{2,n}\|_{P,2} \leq C < \infty$, $M_{2,n} = \max_{1 \leq i \leq n} F_{2,n}(U_i) \leq \frac{C}{\sqrt{\tau_n}}$.

In addition, I have $E G^2_n(\Delta, k) = O\left(\frac{1}{\sqrt{n\tau_n}}\right) = o(1)$. This implies $\sqrt{n}(P_n - P)f \rightsquigarrow 0$ on any subset of $F_{2,n}$ with finite number of elements. On the other hand, the empirical process indexed by $f \in F_{2,n}$ is stochastically equicontinuous. To see this, consider $F_{2,n}^\delta = \{f - g, f, g \in F_{2,n}, \|f - g\|_{P,2} \leq \delta\}$ with an envelope $F_{2,n}^\delta = 2F_{2,n}$ and $M_{2,n}^\delta = \frac{C}{\sqrt{\tau_n}}$. By applying Lemma A.1.1 on $F_{2,n}^\delta$ with $\sigma := \delta$, the Markov inequality,
and the fact that \( \tau_n n \to \infty \), I obtain that for any \( \varepsilon > 0 \),

\[
\lim_{\delta \downarrow 0} \limsup_n P \left( \| \sqrt{n} (P_n - P) \|_{\mathcal{F}_{2,n}} \geq \varepsilon \right) \leq \lim_{\delta \downarrow 0} \limsup_n C \varepsilon^{-1} \left( \sqrt{v \delta^2 \log \left( \frac{2a \| P_{2,n} \|_{P,2}}{\delta} \right)} + \frac{v}{\sqrt{n \tau_n}} \log \left( \frac{2a \| P_{2,n} \|_{P,2}}{\delta} \right) \right) = 0.
\]

This implies \( \sup_{|\Delta| \leq M, k \in [\kappa_1, \kappa_2]} |G_n(\Delta, k) - \mathbb{E}G_n(\Delta, k)| = \sqrt{n} \| P_n - P \|_{\mathcal{F}_{2,n}} = o_p(1) \).

Combining (A.1.8) and (A.1.9), I obtain that

\[
G_n(\Delta, k) := \sqrt{n} P_n f \xrightarrow{p} \frac{\Delta^2}{2}
\]

uniformly in \( \Delta \) and \( k \). Then, combining (C.0.28) and (A.1.10), I obtain (A.1.2). This concludes step 1.

Step 2.

Next I consider \( \hat{W}_n \) in (A.1.1):

\[
\hat{W}_n(k) = J_{n,1}(k) - J_{n,2}(k) + J_{n,3}(k)
\]

where

\[
J_{n,1}(k) := \frac{1}{\sqrt{n k \tau_n}} \sum_{i=1}^{n} \frac{D_i}{P(X_i)} T_{i,1,n}(k),
\]

\[
J_{n,2}(k) := \frac{1}{\sqrt{n k \tau_n}} \sum_{i=1}^{n} \frac{D_i (\hat{P}(X_i) - P(X_i))}{P(X_i)^2} T_{i,1,n}(k),
\]

\[
J_{n,3}(k) := \frac{1}{\sqrt{n k \tau_n}} \sum_{i=1}^{n} \frac{D_i (\hat{P}(X_i) - P(X_i))^2}{P(X_i)^2 P(X_i)} T_{i,1,n}(k),
\]

and \( T_{i,1,n}(k) = k \tau_n - 1 \{ Y_{i,1} \leq q_1(k \tau_n) \} \). Note that \( T_{i,1,n}(k) \) has an envelope

\[
\sup_k |T_{i,1,n}(k)| \leq \overline{T}_{i,1,n} := \kappa_2 \tau_n + 1 \{ Y_{i,1} \leq q_1(\kappa_2 \tau_n) \}.
\]

169
In the following, I will bound \((J_{n,1}(k), J_{n,2}(k), J_{n,3}(k))\) uniformly over \(k \in [\kappa_1, \kappa_2]\).

For \(J_{n,3}(k)\), I have

\[
\sup_k |J_{n,3}(k)| \leq \frac{1}{\sqrt{n\kappa_1 \tau_n}} \sum_{i=1}^{n} |\tilde{T}_{i,1,n}| \varrho_p \left( \frac{1}{\sqrt{n}} \right) = o_p(1). \tag{A.1.11}
\]

This is based on two observations: (1) \(\mathbb{E} \sup_k \sum_{i=1}^{n} |T_{i,1,n}| \leq n \mathbb{E} T_{i,1,n} = C n \tau_n\), so \(\sum_{i=1}^{n} |T_{i,1,n}| = O_p(n \tau_n)\); (2) under Assumption 3, Lemma 1 of Hirano et al. (2003) shows that \(\sup_x |\hat{p}(x) - P(x)| = o_p(n^{-1/4})\).

For \(J_{n,2}(k)\), I have \(J_{n,2}(k) = J_{n,4}(k) + J_{n,5}(k)\) where

\[
J_{n,4}(k) := \sqrt{\frac{n}{k \tau_n}} \int_{\text{Supp}(X)} \frac{1}{P(x)} (\hat{p}(x) - P(x)) (\mathbb{E}(T_{i,1,n}(k) | x)) dF_X(x)
\]

and

\[
J_{n,5}(k) := \frac{1}{\sqrt{n k \tau_n}} \sum_{i=1}^{n} \left[ \frac{D_i}{P(X_i)^2} (\hat{p}(X_i) - P(X_i)) T_{i,1,n}(k) \right. \\
- \left. \int_{\text{Supp}(X)} \frac{1}{P(x)} (\hat{p}(x) - P(x)) (\mathbb{E}(T_{i,1,n}(k) | x)) dF_X(x) \right].
\]

Next, I show \(J_{n,5}(k) = o_p^2(1)\). Denote \(P_h(x) = L(H_h(x)' \pi_h)\) where

\[
\pi_h = \arg \min_{\pi \in \mathbb{R}^h} \mathbb{E}(P(X) \log(L(H_h(X) \pi)) + (1 - P(X)) \log(1 - L(H_h(X)' \pi))),
\]

\(H_h(X)\) is the series bases used for approximation such as polynomials or B-splines, and \(h\) is the number of terms of the series. For \(J_{n,5}(k)\), I have \(J_{n,5}(k) = J_{n,6}(k) + J_{n,7}(k)\) where

\[
J_{n,6}(k) := \frac{1}{\sqrt{n k \tau_n}} \sum_{i=1}^{n} \left( \frac{D_i T_{i,1,n}(k)}{P(X_i)^2} (\hat{p}(X_i) - P_h(X_i)) \right. \\
- \left. \int_{\text{Supp}(X)} \frac{\mathbb{E}(T_{i,1,n}(k) | x)}{P(x)} (\hat{p}(x) - P_h(x)) dF_X(x) \right).
\]
and

\[ J_{n,7}(k) := \frac{1}{\sqrt{n}k_{\tau_n}} \sum_{i=1}^{n} \text{biggl} \left( \frac{D_iT_{i,1,n}(k)}{P(X_i)^2} (P_h(X_i) - P(X_i)) \right) \]

\[- \int_{\text{Supp}(X)} \frac{\mathbb{E}(T_{i,1,n}(k)|x)}{P(x)} (P_h(x) - P(x))dF_X(x) \biggr) \right].

By Lemma 1 of Hirano et al. (2003), \( \sup_x |P_h(x) - P(x)| \leq \zeta(h)h^{-\frac{7}{2}} \) where \( \zeta(h) = \sup_x ||H_h(x)|| \) and \( ||A|| = \sqrt{tr(A^TA)} \). For polynomial bases, \( \zeta(h) \leq Ch \). All the rates restriction in Assumption 3 are stated under this circumstance.

Next, I first compute the order of magnitude of \( J_{n,7}(k) \).

\[ J_{n,7}(k) = \sqrt{n}(\mathcal{P}_n - \mathcal{P}) f, f \in \mathcal{F}_{3,n} \]

where

\[ \mathcal{F}_{3,n} = \left\{ \frac{1}{\tau_n} \left( \frac{D_iT_{i,1,n}(k)}{P(X_i)^2} (P_h(X_i) - P(X_i)) \right) \right. \]

\[ - \int_{\text{Supp}(X)} \frac{\mathbb{E}(T_{i,1,n}(k)|x)}{P(x)} (P_h(x) - P(x))\mathbb{E}(T_{i,1,n}(k)|x)dF_X(x) \biggr), k \in [\kappa_1, \kappa_2] \}

with an envelope function \( F_{3,n} = \frac{C}{\sqrt{n}}(\bar{T}_{i,1,n} + \mathbb{E}(\bar{T}_{i,1,n}|X)) \). Since

\[ \mathbb{E}J_{n,7}^2(k) \leq \frac{1}{\tau_n} \mathbb{E} \left( \frac{D_iT_{i,1,n}(k)}{P(X_i)^2} (P_h(X_i) - P(X_i)) \right)^2 \leq \zeta(h)^2h^{-\frac{7}{2}} \frac{\mathbb{E}T_{i,1,n}^2(k)}{\tau_n} = o(1), \]

\( J_{n,7}(k) \rightsquigarrow 0 \) on any subsets of \( [\kappa_1, \kappa_2] \) with finite elements. I next show that

\( \sqrt{n}(\mathcal{P}_n - \mathcal{P}) f, f \in \mathcal{F}_{3,n} \) is stochastically equicontinuous.

I note that \( \|F_{3,n}\|_{P,2} \leq C < \infty \) and \( M_{3,n} = \max_{1 \leq i \leq n} F_{3,n}(U_i) \leq \frac{C}{\sqrt{n}} \). Therefore,

\[ \mathcal{F}_{3,n}^\delta = \{ f - g, f, g \in \mathcal{F}_{3,n}, ||f - g||_{P,2} \leq \delta \} \]
with an envelope $2F_{3,n}$ and $M_{3,n}^\delta = \frac{C}{\sqrt{n}}$. In addition, $\{T_{i,1,n}(k) : k \in [\kappa_1, \kappa_2]\}$ satisfies the uniform entropy condition because it is a VC-class, and the class of functions $\{\mathbb{E}(T_{i,1,n}(k)|X) : k \in [\kappa_1, \kappa_2]\}$ is generated by taking the conditional expectation which implies that it also satisfies the uniform entropy condition. Therefore, $\mathcal{F}_{3,n}^\delta$ satisfies the uniform entropy condition, that is,

$$\sup_Q N(\varepsilon\|F_{3,n}^\delta\|_Q, \mathcal{F}_{3,n}^\delta, \|\cdot\|_Q) \leq \left(\frac{a}{\varepsilon}\right)^v, \quad \forall \varepsilon \in (0, 1].$$

By applying Lemma A.1.1 on $\mathcal{F}_{3,n}^\delta$ with $\sigma := \delta$ and the Markov inequality, I have

$$\lim_{\delta \downarrow 0} \limsup_n P\left(\|\sqrt{n}(\mathcal{P}_n - P)\|_{\mathcal{F}_{3,n}^\delta} \geq \varepsilon\right) \leq \lim_{\delta \downarrow 0} \sup C \varepsilon^{-1} \left(\sqrt{v\delta^2 \log \left(\frac{2a\|F_{3,n}^\delta\|_{P,2}}{\delta}\right)} + \frac{v}{\sqrt{n}\tau_n} \log \left(\frac{2a\|F_{3,n}^\delta\|_{P,2}}{\delta}\right)\right) = 0.$$ This verifies that $\sqrt{n}(\mathcal{P}_n - P)f, f \in \mathcal{F}_{3,n}$ is stochastically equicontinuous. Combining this with the finite-dimensional convergence, I obtain that $J_{n,7}(k) = o^*_P(1)$.

For $J_{n,6}(k)$, by the Taylor expansion, I have $J_{n,6}(k) = (W_{1,1}(k) + W_{2,1}(k) - W_{3,1}(k)) (\hat{\pi}_h - \pi_h)$, in which

$$W_{1,1}(k) := \frac{1}{\sqrt{n}k\tau_n} \sum_{i=1}^n \left[ \frac{D_i T_{i,1,n}(k)}{P(X_i)^2} L'(H_h^T(X_i) \pi_h) H_h^T(X_i) \right.$$

$$\left. - \int_{\text{Supp}(X)} \frac{\mathbb{E}(T_{i,1,n}(k)|x)}{P(x)} L'(H_h^T(x) \pi_h) H_h^T(x) dF_X(x) \right],$$

$$W_{2,1}(k) := \frac{1}{\sqrt{n}k\tau_n} \sum_{i=1}^n \left[ \frac{D_i T_{i,1,n}(k)}{P(X_i)^2} L''(H_h^T(X_i) \pi_h) H_h(X_i) H_h^T(X_i) (\pi_h - \hat{\pi}_h), \right.$$ and

$$W_{3,1}(k) := \sqrt{\frac{n}{k\tau_n}} \int_{\text{Supp}(X)} \frac{\mathbb{E}(T_{i,1,n}(k)|x)}{P(x)} L''(H_h^T(x) \pi_h) H_h(x) H_h^T(x) (\pi_h - \hat{\pi}_h).$$

172
For an arbitrary deterministic sequence $l_n \to \infty$ and $f \in \mathcal{F}_{4,n}$,

$$
\frac{W_{1,h}(k)}{\zeta(h) l_n} = \sqrt{n}(P_n - \mathcal{P})f
$$

where

$$
\mathcal{F}_{4,n} = \left\{ \frac{1}{\sqrt{\tau_n \zeta(h) l_n}} \left[ \frac{D_i T_{i,1,n}(k)}{P(X_i)} L'(H_h^T(X_i) \pi_h) H_h^T(X_i) \right. \\
- \left. \int_{\text{supp}(x)} \frac{\mathbb{E}(T_{i,1,n}(k)|x)}{P(x)} L'(H_h^T(x) \pi_h) H_h^T(x) dF_X(x) \right], k \in [\kappa_1, \kappa_2] \right\}
$$

with an envelope function

$$
F_{4,n} = \frac{C}{\sqrt{\tau_n \zeta(h) l_n}} (H_h^T(X_i) \overline{T}_{i,1,n} + \int H_h^T(x) \mathbb{E}(\overline{T}_{i,1,n}|X = x) dF_X(x)).
$$

Since

$$
\mathbb{E}\|W_{1,h}(k)\|^2 \lesssim \left( \frac{\mathbb{E}T_{i,1,n}^2(k)}{\tau_n} \right) \zeta^2(h) = O(\zeta^2(h)),
$$

$\{\sqrt{n}(P_n - \mathcal{P})f : f \in \mathcal{F}_{4,n}\} \quad \leadsto \quad 0$ in finite dimension. In addition, $\|F_{4,n}\|_{P,2} \leq C < \infty$ and $M_{4,n} = \max_{1 \leq i \leq n} F_{4,n}(X_i) \leq \frac{C}{\sqrt{\tau_n \ln n}}$. Therefore, for

$$
\mathcal{F}_{4,n}^\delta = \{ f - g, f, g \in \mathcal{F}_{4,n}, \|f - g\|_{P,2} \leq \delta \}
$$

with an envelope $2F_{4,n}$, I have $\|F_{4,n}^\delta\|_{P,2} \leq C$, $M_{4,n}^\delta = \frac{C}{\sqrt{\tau_n \ln n}}$, and

$$
\sup_Q N(\varepsilon\|F_{4,n}^\delta\|_{Q,2}, \mathcal{F}_{4,n}^\delta, \|\cdot\|_{Q,2}) \leq \left( \frac{a}{\varepsilon} \right)^v, \quad \forall \varepsilon \in (0, 1].
$$

By applying Lemma A.1.1 on $\mathcal{F}_{4,n}^\delta$ with $\sigma := \delta$ and the Markov inequality, I have

$$
\lim_{\delta \downarrow 0} \lim_{n} \sup P\left( \left\{ \sqrt{n}(P_n - \mathcal{P}) \right\}_{\mathcal{F}_{4,n}^\delta} \geq \varepsilon \right)
$$

$$
\leq \lim_{\delta \downarrow 0} \lim_{n} C \varepsilon^{-1} \left( \sqrt{v} \delta^2 \log \left( \frac{2a \|F_{4,n}^\delta\|_{P,2}}{\delta} \right) + \frac{v}{\sqrt{\tau_n l_n}} \log \left( \frac{2a \|F_{4,n}^\delta\|_{P,2}}{\delta} \right) \right) = 0.
$$
Therefore, \( W_{1,h}(k) = o_p^*(\zeta(h)l_n) \) for any sequence of \( l_n \) such that \( l_n \to \infty \).

For \( W_{2,h}(k) \),
\[
\mathbb{E} \sup_k ||W_{2,h}(k)|| \leq \mathbb{E} \left[ \frac{D_iT_{i,1,n}L''}{P(X_i)} ||H_h(X_i)||^2 ||\pi_h - \pi_h|| \frac{n}{\sqrt{n\tau_n}} = O(\zeta(h)^2 \sqrt{h}) \right]
\]
So \( W_{2,h}(k) = O_p^*(\zeta(h)^2 \sqrt{h}) \). Similarly,
\[
\mathbb{E} \sup_k ||W_{3,h}(k)|| \\
\leq \sqrt{\frac{n}{\tau_n}} \int_{\text{Supp}(X)} \left[ \frac{\mathbb{E}(T_{i,1,n}|x)}{P(x)} L'' ||H_h(x)||^2 dF_X(x) ||\pi_h - \pi_h|| = O(\zeta(h)^2 \sqrt{h}) \right]
\]
So \( W_{3,h}(k) = O_p^*(\zeta(h)^2 \sqrt{h}) \). Combining all the results, \( J_{n,7}(k) = O_p^*(\zeta(h)^2 \sqrt{h}) \sqrt{\frac{\tau_n}{n}} = o_p^*(1) \) and thus \( J_{n,5}(k) = o_p^*(1) \).

I next decompose \( J_{n,4} \): \( J_{n,4}(k) = J_{n,8}(k) + J_{n,9}(k) \) where
\[
J_{n,8}(k) := \sqrt{\frac{n}{k\tau_n}} \int_{\text{Supp}(X)} \frac{\mathbb{E}(T_{i,n,i}(k)|x)}{P(x)} (\hat{P}(x) - P_h(x))dF_X(x),
\]
\[
J_{n,9}(k) := \sqrt{\frac{n}{k\tau_n}} \int_{\text{Supp}(X)} \frac{\mathbb{E}(T_{i,n,i}(k)|x)}{P(x)} (P_h(x) - P(x))dF_X(x).
\]

For \( J_{n,9}(k) \) I have,
\[
J_{n,9}(k) \leq \sqrt{\frac{n}{k\tau_n}} \int_{\text{Supp}(X)} \left[ \frac{\mathbb{E}(T_{i,n,i}(k)|x)}{P(x)} dF_X(x) \zeta(h)h^{-\frac{\tau}{2}} = O_p^*(\sqrt{n\tau_n} \zeta(h)h^{-\frac{\tau}{2}}) = o_p^*(1). \right.
\]

For \( J_{n,8} \), by the Taylor expansion,
\[
J_{n,8}(k) = \sqrt{\frac{n}{k\tau_n}} \int_{\text{Supp}(X)} \frac{\mathbb{E}(T_{i,n,i}(k)|x)}{P(x)} L'(H_h(x)^T \pi_h) H_h(x)^T dF_X(x)(\hat{\pi}_h - \pi_h).
\]

Since \( \hat{\pi}_h \) solves the first order condition, \( \hat{\pi}_h - \pi_h = \frac{1}{n} \sum_{i=1}^n (\hat{\Sigma}_h)^{-1} (D_i - P_h(X_i))H_h(X_i) \), in which
\[
\hat{\Sigma}_h = \frac{1}{n} \sum_{i=1}^n L'(H_h(X_i)^T \pi_h) H_h(X_i) H_h(X_i)^T.
\]
Hence, I have

\[ J_{n,8}(k) \]

\[ = \frac{1}{n} \sum_{i=1}^{n} \sqrt{\frac{n}{\tau_n}} \int_{\text{Supp}(X)} \frac{\mathbb{E}(T_{i,n,i}|x)}{P(x)} L'(H_h(x)^T \tilde{\pi}_h) H_h(x)^T dF_X(x)(\tilde{\Sigma}_h)^{-1}(D_i - P_h(X_i))H_h(X_i) \]

\[ = \bar{\Psi}_h^T(k)(\tilde{\Sigma}_h)^{-1}V_h \]

\[ = \Psi_h^T(k)\Sigma_h^{-1}V_h + (\tilde{\Psi}_h^T(k) - \Psi_h^T(k))\tilde{\Sigma}_h^{-1}V_h + \Psi_h^T(k)(\tilde{\Sigma}_h^{-1} - \Sigma_h^{-1})V_h \]

\[ := \Psi_h^T(k)\Sigma_h^{-1}V_h + J_{n,10}(k) + J_{n,11}(k) \]

where

\[ \bar{\Psi}_h(k) := \frac{1}{\sqrt{\tau_n}} \int_{\text{Supp}(X)} \frac{\mathbb{E}(T_{i,1,n}|x)}{P(x)} L'(H_h(x)^T \tilde{\pi}_h) H_h(x) dF_X(x), \]

\[ \Psi_h(k) := \frac{1}{\sqrt{\tau_n}} \int_{\text{Supp}(X)} \frac{\mathbb{E}(T_{i,1,n}|x)}{P(x)} L'(H_h(x)^T \pi_h) H_h(x) dF_X(x), \]

\[ \Sigma_h := \mathbb{E}(H_h(x)H_h(x)^T L'(H_h(x)^T \pi_h)), \]

\[ V_h := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} H_h(X_i)(D_i - P_h(X_i)). \]

Since \( \lambda_{\min}(\tilde{\Sigma}_h) \geq \varepsilon > 0 \), \( V_h = O_p(\zeta(h)) \), and

\[ \|\|\| \bar{\Psi}_h(k) - \Psi_h(k) \|\| \| \]

\[ \lesssim \frac{1}{\sqrt{\kappa_1 \tau_n}} \int_{\text{Supp}(X)} \frac{\mathbb{E}(T_{i,n}|x)}{P(x)} L''(H_h(x)^T \tilde{\pi}_h) \|\|H_h(x)\|\|^2 dF_X(x) \|\|\tilde{\pi}_h - \pi_h\|\| \]

\[ = O_p^*(\sqrt{\frac{h}{n}})^2 \sqrt{\frac{h}{n}}, \]

I have \( J_{n,10}(k) = O_p^*(\sqrt{\frac{h}{n}})^3 = o_p^*(1) \).

For \( J_{n,11}(k) \), I first denote \( \hat{\Sigma}_h = \frac{1}{n} \sum_{i=1}^{n} L'(H_h(X_i)^T \pi_h)H_h(X_i)H_h(X_i)^T \). By notic-

\[ 1 \lambda_{\min}(A) \text{ is the minimal eigenvalue of a positive definite matrix } A. \]
ing that $\mathbb{E}||V_h||^2 = O(\zeta(h)^2)$, I have

$$||\hat{\Sigma}_h - \Sigma_h|| \Sigma_h^{-1} V_h||$$

$\leq ||(\hat{\Sigma}_k - \hat{\Sigma}_k)|| \Sigma_h^{-1} V_h|| + ||(\hat{\Sigma}_k - \Sigma_h)|| \Sigma_h^{-1} V_h||$

$\leq \frac{1}{n} \sum_{i=1}^{n} ||H_h(X_i)^T (\hat{\pi}_h - \pi_h) L''(H_h(X_i)^T \pi_h) H_h(X_i) H_h(X_i)^T \Sigma_h^{-1} V_h||$

$+ \frac{1}{n} \sum_{i=1}^{n} [L'(H_h(X_i)^T \pi_h) H_h(X_i) H_h(X_i)^T - \mathbb{E}L'(H_h(X_i)^T \pi_h) H_h(X_i) H_h(X_i)^T] \Sigma_h^{-1} V_h||$

$\leq O_p(\zeta(h)^4 \sqrt{\frac{h}{n}}) + O_p(\frac{1}{\sqrt{n}} \mathbb{E}||L'(H_h(X_i)^T \pi_h) H_h(X_i) H_h(X_i)^T||^2)^{1/2} ||\zeta(h)||$

$= O_p(\zeta(h)^4 \sqrt{\frac{h}{n}} + \zeta(h)^3 \cdot )$.

Furthermore, $||\Psi_h(k)|| \leq O_p(\frac{\zeta(h)^5}{\sqrt{n}} \mathbb{E}(T_{1,1,n}|x))) = O_p(\sqrt{n} \zeta(h))$. This implies

$$J_{n,11}(k) = O_p(\sqrt{n} \zeta(h)^5 \sqrt{\frac{h}{n}} + \zeta(h)^4 \sqrt{n})$$

and

$$J_{n,8}(k) = \Psi^T_h(k) \Sigma_h^{-1} V_h + O_p(\sqrt{n} \zeta(h)^5 \sqrt{\frac{h}{n}}) = \Psi_h(k)^T \Sigma_h^{-1} V_h + o^*_p(1).$$

Next, I compute the leading term of $J_{n,8}(k)$: $\Psi^T_h(k) \Sigma_h^{-1} V_h$. Define

$$\delta_0(x, k) := \frac{\mathbb{E}(T_{1,1,n}(k)|x)}{\sqrt{k \tau_n} P(x)} \sqrt{P(x)(1 - P(x))},$$

$$\delta_h(x, k) := \Psi^T_h(k) \Sigma_h^{-1} \sqrt{P_h(x)(1 - P_h(x))} H_h(x).$$

Then

$$\Psi^T_h(k) \Sigma_h^{-1} V_h = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_h(X_i, k) - D_i - P_h(X_i) \sqrt{P_h(X_i)(1 - P_h(X_i))}.$$
I want to compute the difference

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \delta_h(X_i, k) \frac{D_i - P_h(X_i)}{\sqrt{P_h(X_i)(1 - P(X_i))}} - \delta_0(X_i) \frac{D_i - P(X_i)}{P(X_i)(1 - P(X_i))} \right] = J_{n,12}(k) + J_{n,13}(k)
\]

where

\[
J_{n,12}(k) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ (\delta_h(X_i, k) - \delta_0(X)) \frac{D - P(X_i)}{P(X_i)(1 - P(X_i))} \right],
\]

\[
J_{n,13}(k) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \delta_h(X_i, k) \left( \frac{D_i - P_h(X_i)}{\sqrt{P_h(X_i)(1 - P(X_i))}} - \frac{(D_i - P(X_i))}{\sqrt{P(X_i)(1 - P(X_i))}} \right) \right].
\]

For \( J_{n,12}(k) \), notice that \( \sqrt{\tau_n} \delta_h(x, k) \) is the projection of

\[
\sqrt{\tau_n} \delta_0(x, k)
\]

on

\[
\sqrt{\mathcal{L}(H_h(x) \tau_h)} H_h(x).
\]

By Assumption 3, \( \mathbb{E}(T_{i,1,n}(k)|x) \) and \( P(x) \) are \( t \) times differentiable with their derivatives being bounded by \( M_n \) on Supp(\( X \)) uniformly over the quantile index (and thus \( k \)). Hence,

\[
\sup_{(x,k) \in \text{Supp}(X) \times [s_1,s_2]} \| \delta_0(x, k) - \delta_h(x, k) \| \leq M_n h^{-\frac{t}{2}} / \sqrt{\tau_n}
\]

and

\[
J_{n,12}(k) = O_p^* \left( \sqrt{\frac{nM_n}{\tau_n} h^{-\frac{t}{2}}} \right) = o_p^*(1).
\]

For \( J_{n,13}(k) \), I have

\[
\| J_{n,13}(k) \| \leq \sqrt{n} \sup_{k,x} \| \delta_h(x, k) \| \zeta(h) h^{-\frac{t}{2}} = O_p^* \left( \sqrt{n\tau_n} \zeta^3(h) h^{-\frac{t}{2}} \right) = o_p^*(1).
\]
Combining the bounds on \((J_{n,10}(k), \cdots, J_{n,13}(k))\), I obtain that

\[
J_{n,8}(k) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_h(X_i, k) \frac{D_i - P_h(X_i)}{\sqrt{P_h(X_i)(1 - P_h(X_i))}} + o_p^*(1)
\]

\[
= \frac{1}{\sqrt{n}k\tau_n} \sum_{i=1}^{n} \frac{\mathbb{E}(T_{i,1,n}(k)|X_i)}{P(X_i)} (D_i - P(X_i)) + o_p^*(1).
\]

Then by combining \(J_{n,1}(k) - J_{n,8}(k)\), I have,

\[
\hat{W}_n(k) = W_n(k) + o_p^*(1).
\]

This concludes Step 2.

Step 3.

Note that

\[
W_n(k) = \sqrt{n}(\mathcal{P}_n - \mathcal{P})f
\]

for \(f \in \mathcal{F}_{5,n}\), in which \(\mathcal{F}_{5,n} = \{\phi_{i,1,n}(k), k \in [\kappa_1, \kappa_2]\}\) and

\[
\phi_{i,1,n}(k) = \frac{1}{\sqrt{k\tau_n}} \left[ \frac{D_i}{P(X_i)} T_{i,1,n}(k) - \frac{\mathbb{E}(T_{i,1,n}(k)|X_i)}{P(X_i)} (D_i - P(X_i)) \right].
\]

Then,

\[
F_{5,n} = \frac{C}{\sqrt{\tau_n}} (T_{i,1,n}|X_i + \mathbb{E}(T_{i,1,n}|X_i))
\]

is an envelope for \(\mathcal{F}_{5,n}\). We have \(||F_{5,n}||_{P,2} \leq C < \infty\). \(M_{5,n} := \max_{1 \leq i \leq n} F_{5,n}(Y_i, X_i) \leq \frac{C}{\sqrt{\tau_n}}\).

First notice that, for \(f \in \mathcal{F}_{5,n}\), \(\mathcal{P}f = 0\), \(Pf^2 \leq \frac{1}{\tau_n} \mathbb{E}T_{i,1,n}^2(k) = O(1)\). So the empirical process \(\sqrt{n}(\mathcal{P}_n - \mathcal{P})f\) indexed by \(f \in \mathcal{F}_{5,n}\) is bounded in probability in any subsets of \(\mathcal{F}_{5,n}\) with finite number of elements.

Next, I want to show the empirical process is stochastically equicontinuous. Let

\[
\mathcal{F}_{5,n}^\delta = \{f - g, f, g \in \mathcal{F}_{5,n}, ||f - g||_{P,2} \leq \delta\}
\]

178
with envelope $2F_{5,n}$. Then similar to $\mathcal{F}_{\delta,n}$, there exists $v > 0$ and $a > e$ such that

$$
\sup_Q N(\varepsilon \|F_{5,n}\| Q,2, \mathcal{F}_{\delta,n}^{\delta}, \|\| Q,2) \leq \left(\frac{a}{\varepsilon}\right)^v , \quad \forall \varepsilon \in (0, 1].
$$

By applying Lemma A.1.1 on $\mathcal{F}_{\delta,n}$ with $\sigma := \delta$ and the Markov inequality, I have

$$
\lim_{\delta \downarrow 0} \limsup_n P \left( \|\sqrt{n}(P_n - P)\|_{\mathcal{F}_{\delta,n}^{\delta}} \geq \varepsilon \right) 
\leq \lim_{\delta \downarrow 0} \limsup_n C\varepsilon^{-1} \left( \sqrt{n} \delta^2 \log \left( \frac{2a \|F_{5,n}\|_{P,2}}{\delta} \right) + \frac{v}{\sqrt{n} \tau_n} \log \left( \frac{2a \|F_{5,n}\|_{P,2}}{\delta} \right) \right) = 0.
$$

Therefore, the empirical process $\sqrt{n}(P_n - P)$ indexed by $f \in \mathcal{F}_{5,n}$ is stochastically equicontinuous and the stochastic process $\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_{i,1,n}(k) : k \in [\kappa_1, \kappa_2] \right\}$ is tight.

It further implies the stochastic process $\{ \hat{W}_n(k) : k \in [\kappa_1, \kappa_2] \}$ is tight. This concludes Step 3 as well as the proof of the first part of Theorem 2.3.1.

I next turn to the proof of the second part of Theorem 2.3.1. By the additional assumption in the theorem, the covariance kernel satisfies that

$$
\mathbb{E}(\phi_{i,1,n}(k_1), \phi_{i,0,n}(k_2))(\phi_{i,1,n}(k_1), \phi_{i,0,n}(k_2))' \rightarrow \mathcal{H}(k_1, k_2).
$$

This is sufficient for the finite-dimensional convergence of

$$
(\lambda_{1,n}(k)(\hat{q}_1(k\tau_n) - q_1(k\tau_n)), \lambda_{0,n}(k)(\hat{q}_0(k\tau_n) - q_0(k\tau_n))).
$$

Combining the finite-dimensional convergence with the stochastic equicontinuity of

$$
\left\{ (\lambda_{1,n}(k)(\hat{q}_1(k\tau_n) - q_1(k\tau_n)), \lambda_{0,n}(k)(\hat{q}_0(k\tau_n) - q_0(k\tau_n))), k \in [\kappa_1, \kappa_2] \right\},
$$

I have shown the second part of Theorem 2.3.1.
A.1.5 Proof of Theorem 2.3.3

Hereafter, all bootstrap counterparts are starred. Let \( \{I_{n,j}\}_{j \geq 1} \) denote an i.i.d. sequence distributed as multinomial with parameter \( 1 \) and probability \( \left( \frac{1}{n}, \ldots, \frac{1}{n} \right) \), so that the bootstrap weight for individual \( i \), \( w_{n,i} \), satisfies \( w_{n,i} = \sum_{j=1}^{n} 1\{I_{n,j} = i\} \). Also, let \( \hat{\Delta}_{1,n}^* = \lambda_{1,n}(\hat{q}_1^*(\tau_n) - q(\tau_n)) \) where \( \lambda_{1,n} \) is defined in (2.3.3). Similar to the proof of Theorem 2.3.1,

\[
\hat{\Delta}_{1,n}^* = \arg \min_{\Delta \in \mathbb{R}} -\hat{W}_n^* \Delta + \hat{G}_n^*(\Delta)
\]  

(A.1.12)

where

\[
\hat{W}_n^* = \frac{1}{\sqrt{n\tau_n}} \sum_{i=1}^{n} \frac{w_{n,i} D_i}{P(X_i)} (\tau_n - 1\{Y_i \leq q_1(\tau_n)\}),
\]

\[
\hat{G}_n^*(\Delta) = \frac{1}{\sqrt{n\tau_n}} \sum_{i=1}^{n} \frac{w_{n,i} D_i}{P(X_i)} \int_0^\Delta \left( 1\left\{Y_i \leq q_1(\tau_n) + s \frac{s}{\lambda_{1,n}} \right\} - 1\{Y_i \leq q_1(\tau_n)\} \right) ds.
\]

Since \( Ew_{n,i} = 1 \), same as in the proof of Theorem 2.3.1,

\[
\hat{G}_n^*(\Delta) = \frac{\Delta^2}{2} + o_p(1).
\]  

(A.1.13)

Next, let \( w_{N,n,i} = \sum_{j=1}^{N_n} I_{n,j} = i \), so that \( \{w_{N,n,i}\}_{i=1}^{n} \) are i.i.d. Poisson random variable with unit mean. Let

\[
\tilde{W}_n^* = \frac{1}{\sqrt{n\tau_n}} \sum_{i=1}^{n} \frac{w_{N,n,i} D_i}{P(X_i)} (\tau_n - 1\{Y_i \leq q_1(\tau_n)\}).
\]

I aim to show that

\[
\hat{W}_n^* - \tilde{W}_n^* = o_p(1).
\]

Fix \( \eta > 0 \) and let \( \mathcal{I}_j = \{i : |w_{N,n,i} - w_{n,i}| \geq j\} \) and \( n_j = \#\mathcal{I}_j \). Then, for \( n \) large enough and with a probability greater than \( 1 - \eta \) (see (Van der Vaart and Wellner, 180...
\[
\hat{W}_n - \tilde{W}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (w_{N,i} - w_{n,i}) M_{n,i}(\tau_n) = \text{sign}(N_n - n) \sum_{j=1}^{2} \frac{1}{\sqrt{n}} \sum_{i \in I_j} M_{n,i}(\tau_n)
\]

(A.1.14)

with 
\[
M_{n,i}(\tau_n) = \frac{1}{\sqrt{\tau_n}} \frac{D_i}{P(X_i)} (\tau_n - 1\{Y_i \leq q_1(\tau_n)\}) \quad \text{and the convention that} \quad \sum_{i \in I_j} M_{n,i}(\tau_n) = 0 \quad \text{when} \quad n_j = 0.
\]

I now show that 
\[
\sum_{i \in I_j} M_{n,i}(\tau_n)/\sqrt{n} = o_p(1).
\]

Note that
\[
M_{n,i}(\tau_n) = M_{n,i}^*(\tau_n) + R_{n,i}
\]

where
\[
M_{n,i}^*(\tau_n) = \frac{1}{\sqrt{\tau_n}} \frac{D_i}{P(X_i)} (\tau_n - 1\{Y_i \leq q_1(\tau_n)\})
\]

and
\[
R_{n,i} = \frac{1}{\sqrt{\tau_n}} \frac{D_i(P(X_i) - \hat{P}(X_i))}{P(X_i)P(X_i)} (\tau_n - 1\{Y_i \leq q_1(\tau_n)\}).
\]

I first show
\[
\sum_{i \in I_j} R_{n,i}/\sqrt{n} = o_p(1).
\]

(A.1.15)

Note that
\[
\left| \sum_{i \in I_j} R_{n,i}/\sqrt{n} \right| \leq \frac{1}{\sqrt{n\tau_n}} \sum_{i \in I_j} |\tau_n - 1\{Y_{i,1} \leq q_1(\tau_n)\}| \sup_{x \in \text{Supp}(X)} |\hat{P}(x) - P(x)|
\]

\[
\leq \frac{1}{\sqrt{n\tau_n}} \sum_{i \in I_j} |\tau_n - 1\{Y_{i,1} \leq q_1(\tau_n)\}| o_p(1).
\]

In addition,
\[
\frac{1}{n\tau_n} E \left[ \left( \sum_{i \in I_j} |\tau_n - 1\{Y_{i,1} \leq q_1(\tau_n)\}| \right)^2 \right] \left( I_{n,j} \right)_{j \geq 1}, N_n \right] \leq \left( \frac{n_j}{\sqrt{n}} \right)^2 \leq \left( \frac{N_n - n}{\sqrt{n}} \right)^2 = O_p(1).
\]

181
Thus (A.1.15) holds. Next, since $E(M^*_{n,i}(\tau_n) | (I_{n,j})_{j \geq 1}, N_n) = 0$ and

$$\frac{1}{n} \text{Var} \left[ \left( \sum_{i \in I_j} M^*_{n,i}(\tau_n) \right) | (I_{n,j})_{j \geq 1}, N_n \right] \leq \frac{n_j}{n} \leq \frac{|N_n - n|}{n} = o_p(1),$$

I have

$$\sum_{i \in I_j} M^*_{n,i}(\tau_n) / \sqrt{n} = o_p(1). \quad (A.1.16)$$

Combining (A.1.15) and (A.1.16), I have shown that $\sum_{i \in I_j} M_{n,i}(\tau_n) / \sqrt{n} = o_p(1)$ and thus

$$\hat{W}_n^* - \tilde{W}_n^* = o_p(1). \quad (A.1.17)$$

In addition, by the same argument in the proof of Theorem 2.3.1, I have

$$\tilde{W}_n^* = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_{N_n,i} \phi_{1,i,n}(1) + o_p(1). \quad (A.1.18)$$

Combining (A.1.12), (A.1.17), and (A.1.18), I obtain that

$$-\hat{W}_n^* \Delta + \hat{G}_n^*(\Delta) = -\left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_{N_n,i} \phi_{1,i,n}(1) \right) \Delta + \frac{\Delta^2}{2}. $$

By the Convexity lemma in Pollard (1991a), I have

$$\hat{\Delta}_{1,n}^* = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_{N_n,i} \phi_{1,i,n}(1) + o_p(1).$$

Recall that, from the proof of Theorem 2.3.1, I have

$$\hat{\Delta}_{1,n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{1,i,n}(1) + o_p(1).$$

Thus

$$\lambda_{n,1}(\hat{q}_1^*(\tau_n) - \hat{q}_1(\tau_n)) = \hat{\Delta}_{1,n}^* - \hat{\Delta}_{1,n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (w_{N_n,i} - 1) \phi_{1,i,n}(1) + o_p(1). \quad (A.1.19)$$
Similarly,

$$\lambda_{n,0}(\hat{q}_0^*(\tau_n) - \hat{q}_0(\tau_n)) = \hat{\Delta}_{0,n}^* - \hat{\Delta}_{0,n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (w_{N,i} - 1) \phi_{i,0,n}(1) + o_p(1). \quad (A.1.20)$$

Also note that, with $C_1(\rho, m)$, $C_0(\rho, m)$, $\lambda_n$, and $\Sigma_n$ defined in Theorem 2.3.2, I have

$$\Sigma_n^{-1/2} \hat{\lambda}_n(\hat{q}(\tau_n) - q(\tau_n)) = \Sigma_n^{-1/2} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (C_1(\rho, m) \phi_{i,1,n}(1) - C_0(\rho, m) \phi_{i,0,n}(1)) + o_p(1) \rightsquigarrow \mathcal{N}(0, 1). \quad (A.1.21)$$

Then combining (A.1.19) and (A.1.19) with the continuous mapping theorem, I obtain that

$$\Sigma_n^{-1/2} \hat{\lambda}_n(\hat{q}^*(\tau_n) - \hat{q}(\tau_n)) = \Sigma_n^{-1/2} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (w_{N,i} - 1) (C_1(\rho, m) \phi_{i,1,n}(1) - C_0(\rho, m) \phi_{i,0,n}(1)) + o_p(1) \rightsquigarrow \mathcal{N}(0, 1).$$

Here the variance $\Sigma_n$ is the same in (A.1.21) because $w_{N,i}$ is independent of data and has unit mean and variance. This concludes the proof.

A.1.6 Proof of Theorem 2.3.4

Note that

$$\frac{\hat{q}_j(ml^r \tau_n) - \hat{q}_j(l^r \tau_n)}{\hat{q}_j(ml^r-1 \tau_n) - \hat{q}_j(l^{r-1} \tau_n)} \sim (1 + O_p(\frac{1}{\sqrt{n}})) \frac{q_j(ml^r \tau_n) - q_j(l^r \tau_n)}{q_j(ml^{r-1} \tau_n) - q_j(l^{r-1} \tau_n)} \sim (1 + O_p(\frac{1}{\sqrt{n}})) l^{-\xi_j}. $$

This implies (1) by the continuous mapping theorem. (2) follows from the delta-method and a triangular array CLT in such as Theorem 3.4.5 in Durrett (2010).
A.1.7 Proof of Theorem 2.4.1

Note that

\[
\hat{Z}_{1,n}(k) = \arg\min_z \frac{1}{\alpha_{1,n}} \left[ -\sum_{i=1}^{n} \frac{D_i}{P(X_i)} (\tau_n z - (z - \alpha_{1,n}(U_{i,1} - \beta_{1,n}))\mathbb{1}\{\alpha_{1,n}(U_{i,1} - \beta_{1,n}) \leq z\}) + \sum_{i=1}^{n} \frac{D_i}{P(X_i)} \tau_n \alpha_{1,n}(U_{i,1} - \beta_{1,n}) \right].
\]

Multiplying the LHS by \(\alpha_{1,n}\) and subtracting

\[
\sum_{i=1}^{n} \frac{D_i}{P(X_i)} (\tau_n \alpha_{1,n}(U_{i,1} - \beta_{1,n}) + (-\delta - \alpha_{1,n}(U_{i,1} - \beta_{1,n}))\mathbb{1}\{\alpha_{1,n}(U_{i,1} - \beta_{1,n}) \leq -\delta\}),
\]

I obtain

\[
\hat{Z}_{1,n}(k) = \arg\min_{z_1} -\sum_{i=1}^{n} W_1(D_i, \hat{P}(X_i)) \tau_n z_1 + \sum_{i=1}^{n} W_1(D_i, \hat{P}(X_i)) l_\delta(\alpha_{1,n}(U_{i,1} - \beta_{1,n}), z_1).
\]

Similarly,

\[
\hat{Z}_{0,n}(k) = \arg\min_{z_0} -\sum_{i=1}^{n} W_0(D_i, \hat{P}(X_i)) \tau_n z_0 + \sum_{i=1}^{n} W_0(D_i, \hat{P}(X_i)) l_\delta(\alpha_{0,n}(U_{i,0} - \beta_{0,n}), z_0).
\]

So overall,

\[
(\hat{Z}_{1,n}(k), \hat{Z}_{0,n}(k)) := \arg\min_{z_1, z_0} \sum_{j=0,1} Q_{j,n}(z_j),
\]

where

\[
Q_{j,n}(z_j) = -\sum_{i=1}^{n} W_j(D_i, \hat{P}(X_i)) \tau_n z_j + \sum_{i=1}^{n} W_j(D_i, \hat{P}(X_i)) l_\delta(\alpha_{j,n}(U_{i,j} - \beta_{j,n}), z_j).
\]

In the following, I divide the proof into five steps. In the first step, I show the marginal convergence, that is, for \(j = 0, 1\) and fixed \(z_j\),

\[
Q_{j,n}(z_j) \rightsquigarrow Q_{j,\infty}(z_j),
\]
in which
\[ Q_{j,\infty}(z_j) = -k z_j + \sum_{i=1}^{\infty} W_j(D_{i,j}, P(X_{i,j})) l_\delta(J_{i,j}, z_j). \]

In the second step, I show that for any \((z_1, z_0), Q_{1,n}(z_1)\) and \(Q_{0,n}(z_0)\) are asymptotically independent. The marginal convergence is sufficient for joint convergences of \((Q_{1,n}(z_1), Q_{0,n}(z_0))\) to \((Q_{1,\infty}(z_1), Q_{0,\infty}(z_0))\). Then by the continuous mapping theorem,
\[ Q_{1,n}(z_1) + Q_{0,n}(z_0) \rightsquigarrow Q_{1,\infty}(z_1) + Q_{0,\infty}(z_0). \]

In the third step, I apply the convexity lemma to show the weak convergence of the sample minimizers \((\hat{Z}_{1,n}(k), \hat{Z}_{0,n}(k))\) to their population counterparts \((Z_{1,\infty}(k), Z_{0,\infty}(k))\) when this \(k\) satisfies Assumption 9.

In the fourth step, I enhance the result to the finite-dimensional convergence, that is, \(k_l\) satisfying Assumption 9,
\[ (\hat{Z}_{1,n}(k_l), \hat{Z}_{0,n}(k_l))_{l=1}^{L} \rightsquigarrow (Z_{1,\infty}(k_l), Z_{0,\infty}(k_l))_{l=1}^{L} \]

\[ := \arg \min_{(z_{1,1}, \ldots, z_{1,1})_{l=1}^{L}} \sum_{j=0,1} \sum_{l=1}^{L} \left\{ -k_l z_{j,l} + \sum_{i=1}^{\infty} W_j(D_{i,j}, P(X_{i,j})) l_\delta(J_{i,j}, z_{j,l}) \right\}. \]

In the last step, I show \((\hat{Z}_{1,n}(k), \hat{Z}_{0,n}(k))\) as a two-dimensional stochastic process indexed by \(k\) in \(D^2([\kappa_1, \kappa_2])\) weakly converges to a two-dimensional stochastic process \((Z_{1,\infty}(k), Z_{0,\infty}(k))\).

Before showing the five steps, I first present four technical statements. Their proofs can be found at the end of the appendix.

**Lemma A.1.2.** Under the conditions of Theorem 2.4.1, for \(j = 0, 1\),
\[ (1) \frac{1}{n} \sum_{i=1}^{n} W_j(D_i, P(X_i)) \to 1 \text{ a.s.} \]
(2) Let

for type 1 tails ($\xi_1 = 0$): $E_j = E^1 = [-\infty, +\infty) \times \{0, 1\} \times \text{Supp}(X)$,

for type 2 tails ($\xi_1 > 0$): $E_j = E^2 = [-\infty, 0) \times \{0, 1\} \times \text{Supp}(X)$,

for type 3 tails ($\xi_1 < 0$): $E_j = E^3 = [0, +\infty) \times \{0, 1\} \times \text{Supp}(X)$.

Then $\hat{N}_j := \sum_{i=1}^n 1\{\alpha_{j,n}(U_{i,j} - \beta_{j,n}), D_i, X_i\}$ as a point process on state space $E_j$
weakly converges to $N_j = \sum_{i=1}^\infty 1\{J_{i,j}, D_{i,j}, X_{i,j}\}$.

(3) Let

\[ g_1(u, x) = \frac{1}{P(x)} l_\delta(u, x, z_1), \quad g_0(u, x) = \frac{1}{1 - P(x)} l_\delta(u, x, z_0), \]

and

\[ \Psi_{j,n} = \sum_{i=1}^n (jD_i + (1 - j)(1 - D_i))g_j(\alpha_{j,n}(U_{i,j} - \beta_{j,n}), X_i). \]

Then for a pair of constants $(t_1, t_0)$, and $\tilde{i}$ representing the imaginary number,

\[ \mathbb{E} \exp \left( it_1 \Psi_{1,n} + \tilde{i} t_0 \Psi_{0,n} \right) \rightarrow \mathbb{E} \exp \left( \tilde{i} \int_{E_1} t_1 dg_1 dN_1 \right) \mathbb{E} \exp \left( \tilde{i} \int_{E_0} t_0 (1 - d) g_0 dN_0 \right), \]

in which $N_j$ is defined in (2).

(4) The distances between two closest discontinuities of the sample paths of the two
marginal stochastic processes $\hat{Z}_{1,n}(k)$ and $\hat{Z}_{0,n}(k)$ indexed by $k$ are both greater than
1.

Step 1:

I focus on the case for $j = 1$ because the case for $j = 0$ can be proved in a similar
manner. First note that for fixed $z_1$, by Lemma A.1.2, $-\sum_{i=1}^n \frac{D_i}{P(X_i)} \tau_n z_1 = -k z_1 +$
\(o_p(1)\). In order to compute the second piece of the objective function, I first define

\[
\theta_{n,1}(z_1) := \sum_{i=1}^{n} \frac{D_i}{P(X_i)} l_\delta(\alpha_{1,n}(U_{i,1} - \beta_{1,n}), z_1),
\]

\[
\theta_{n,2}(z_1) := \sum_{i=1}^{n} \frac{D_i}{P(X_i)} |l_\delta(\alpha_{1,n}(U_{i,1} - \beta_{1,n}), z_1)|,
\]

\[
\theta_{n,3}(z_1) := \sum_{i=1}^{n} \frac{D_i(\hat{P}(X_i) - P(X_i))}{P(X_i) P(X_i)} l_\delta(\alpha_{1,n}(U_{i,1} - \beta_{1,n}), z_1).
\]

Then

\[
\sum_{i=1}^{n} \frac{D_i}{P(X_i)} l_\delta(\alpha_{1,n}(U_{i,1} - \beta_{1,n}), z_1) = \theta_{n,1}(z_1) + \theta_{n,3}(z_1)
\]

and

\[
|\theta_{n,3}(z_1)| \leq \theta_{n,2}(z) \sup_x |\hat{P}(x) - P(x)|.
\]

Also notice that \(\theta_{n,1}(z_1)\) and \(\theta_{n,2}(z_1)\) can be rewritten as

\[
\theta_{n,1}(z_1) = \int_E \frac{d}{P(x)} l_\delta(u, z_1) d\hat{N}_1,
\]

\[
\theta_{n,2}(z_1) = \int_E \frac{d}{P(x)} |l_\delta(u, z_1)| d\hat{N}_1,
\]

in which \(\hat{N}_1\) is defined in Lemma A.1.2. Following part 2 of the proof of Theorem 4.1 in Chernozhukov (2005a), for type 1 and 3 tails, \(\frac{d}{P(x)} l_\delta(u, z_1) \in C_K(E)\) for any fixed \(z\), and for type 2 tails, \(\frac{d}{P(x)} l_\delta(u, z_1) \in C_K(E)\) for \(z_1 < 0\). Also, by Lemma A.1.2(2), \(\hat{N}_1 \rightsquigarrow N_1\). Therefore, for any \(z\) for type 1 and 3 tails and negative \(z\) for type 2 tails,

\[
\theta_{n,1}(z) \rightsquigarrow \theta_{\infty,1}(z_1) = \int_E \frac{d}{P(x)} l_\delta(u, z_1) dN_1
\]

\[
\theta_{n,2}(z) \rightsquigarrow \theta_{\infty,2}(z_1) = \int_E \frac{d}{P(x)} |l_\delta(u, z_1)| dN_1.
\]
This implies that, for the aforementioned region of \( z_1, \theta_{\alpha,2}(z_1) = O_p(1), \theta_{n,3}(z_1) = O_p(\theta_{n,2}(z_1) \sup_x |\hat{P}(x) - P(x)|) = o_p(1) \) and thus

\[
\sum_{i=1}^{n} \frac{D_i}{P(X_i)} l_\delta(\alpha_{1,n}(U_{i,1} - \beta_{1,n}), z_1) \to \theta_{\alpha,1}(z_1).
\]

The last thing to check is \( \sum_{i=1}^{n} \frac{D_i}{P(X_i)} l_\delta(\alpha_{1,n}(U_{i,1} - \beta_{1,n}), z_1) \to +\infty \) for type 2 tails and \( z > 0 \). Again, following Chernozhukov (2005a), under this case, \( \alpha_{1,n} \to 0, \beta_{1,n} = 0, l_\delta(u, z_1) \geq 1 \{-\delta \leq u \leq 0\} z_1 \) if \( u > -\delta \) and \( l_\delta(u, z) = z + \delta \) if \( u \leq -\delta \). Because \( P(\alpha_n U_{i,1} > -\delta) \to 1 \), I have,

\[
\sum_{i=1}^{n} \frac{D_i}{P(X_i)} l_\delta(\alpha_{n} U_{i,1}, z_1) 1\{\alpha_n U_{i,1} \leq -\delta\} \leq \sum_{i=1}^{n} 1\{\alpha_n U_{i,1} \leq -\delta\} = O_p(1),
\]

and

\[
\sum_{i=1}^{n} \frac{D_i}{P(X_i)} l_\delta(\alpha_{n} U_{i,1}, z_1) 1\{\alpha_n U_{i,1} > -\delta\} \geq \sum_{i=1}^{n} 1\{\alpha_n U_{i,1} > -\delta\} = +\infty,
\]

which lead to the desired result that

\[
\sum_{i=1}^{n} \frac{D_i}{P(X_i)} l_\delta(\alpha_{1,n}(U_{i,1} - \beta_{1,n}), z_1) \to +\infty.
\]

By letting \( Q_{1,\alpha}(z_1) = -kz_1 + \int_{E} \frac{d}{\hat{P}(x)} l_\delta(u, z_1) dN_1 \), I have shown that, for all types for tails,

\[
Q_{1,n}(z_1) = Q_{1,\alpha}(z_1).
\]

Similarly, by denoting \( Q_{0,\alpha}(z_0) = -kz_0 + \int_{E} \frac{1 - d}{1 - \hat{P}(x)} l_\delta(u, z_0) dN_0 \), I can show that

\[
Q_{0,n}(z_0) = Q_{0,\alpha}(z_0).
\]

Step 2:

From the proof of step 1, it is sufficient to show the asymptotic independence of

\[
\Psi_{1,n} := \sum_{i=1}^{n} W_1(D_i, P(X_i)) l_\delta(\alpha_{1,n}(U_{i,1} - \beta_{1,n}), z_1)
\]

188
and
\[
\Psi_{0,n} := \sum_{i=1}^{n} W_0(D_i, P(X_i)) l_\delta(\alpha_{0,n}(U_{i,0} - \beta_{0,n}), z_0)
\]
for any \((z_1, z_0)\). Also I have already shown in step 1 that
\[
\Psi_{1,n} \rightsquigarrow \int_{E_1} d g_1(j, d, x) dN_1(j, d, x)
\]
and
\[
\Psi_{0,n} \rightsquigarrow \int_{E_0} (1 - d) g_0(j, d, x) dN_0(j, d, x).
\]
Therefore, I only have to show that, for any pair of constants \((t_1, t_0)\),
\[
\mathbb{E} \exp(it_1 \Psi_{1,n} + it_0 \Psi_{0,n}) \to \mathbb{E} \exp(\int_{E_1} t_1 d g_1 dN_1) \mathbb{E} \exp(\int_{E_0} t_0 (1 - d) g_0 dN_0).
\]
This is done by Lemma A.1.2(3).

Step 3:
From the results in step 1 and 2, I obtain the joint convergence as follows:
\[
(\mathcal{Q}_{1,n}(z_1), \mathcal{Q}_{0,n}(z_0)) \rightsquigarrow (\mathcal{Q}_{1,\infty}(z_1), \mathcal{Q}_{0,\infty}(z_0)), \mathcal{Q}_{1,\infty}(z_1) \perp \mathcal{Q}_{0,\infty}(z_0).
\]
By the continuous mapping theorem,
\[
\mathcal{Q}_{1,n}(z_1) + \mathcal{Q}_{0,n}(z_0) \rightsquigarrow \mathcal{Q}_{1,\infty}(z_1) + \mathcal{Q}_{0,\infty}(z_0).
\]
This result can be easily improved to hold over finite pairs of \((z_1, z_0)\). For fixed \(k\) as the limiting of \(\tau_n n\) who satisfies Assumption 9, I denote
\[
\mathcal{Q}_{j,n}(z_{j,t}) = \sum_{i=1}^{n} \left( - \sum_{i=1}^{n} W_j(D_i, \hat{P}(X_i)) \tau_n z_{j,t} + \sum_{i=1}^{n} W_j(D_i, \hat{P}(X_i)) l_\delta(\alpha_{j,n}(U_{i,j} - \beta_{j,n}), z_{j,t}) \right),
\]
and
\[
\mathcal{Q}_{j,\infty}(z_{j,t}) = \sum_{i=1}^{L} \left( - k z_{j,t} + \sum_{i=1}^{\infty} W_j(D_{i,j}, P(X_{i,j})) l_\delta(J_{i,j}, z_{j,t}) \right).
\]
Then
\[ \sum_{i=1}^{L} [Q_{1,n}(z_{1,i}) + Q_{0,n}(z_{0,i})] \sim \sim \sum_{i=1}^{L} [Q_{1,\infty}(z_{1,i}) + Q_{0,\infty}(z_{0,i})]. \]

This is the finite-dimensional convergence of the objective function. Also notice that \(Q_{1,\infty}(z_{1}) + Q_{0,\infty}(z_{0})\) is convex in \((z_{1}, z_{0})\). Therefore, in order to apply the convexity lemma as in Chernozhukov (2005a), I only have to further verify two statements: (1) \(Q_{j,\infty}(z_{j})\) is finite over a non-empty open set of \((z_{j})\) and (2) \(Z_{j,\infty}(k)\) \(j = 0, 1\) is a unique pair of random variables who minimizes \(\sum_{j=0,1} Q_{j,\infty}(z_{j})\). In fact, (1) can be proved similar to the proof of Theorem 4.1 Part 2(II) in Chernozhukov (2005a). (2) holds by the fact that \(k\) satisfies Assumption 9. One sufficient condition for Assumption 9 is \(k \in [\kappa_{1}, \kappa_{2}]/(L_{1} \cup L_{2})\), in which

\[ L_{j} = \left\{ k \in [\kappa_{1}, \kappa_{2}] : P \left( \sum \frac{1}{P(A_{i,j})} = k \right) > 0 \text{ or } P \left( \sum \frac{1}{P(A_{i,j})} + \frac{1}{P(A_{j,h})} = k \right) > 0 \right\}. \]

for some \(h\) and \(\mu \in \mathcal{M}(l), l \leq h - 1\).

Lemma A.1.5 and A.1.6 show that when \(k \in [\kappa_{1}, \kappa_{2}]/(L_{1} \cup L_{2})\), uniqueness and tightness of \(Z_{j,\infty}(k), j = 0, 1\) hold. This sufficient condition will be used later in the proof.

Then, the convexity lemma implies that
\[ (\hat{Z}_{1,n}(k), \hat{Z}_{0,n}(k)) \sim \sim (Z_{1,\infty}(k), Z_{0,\infty}(k)) \]
\[ := \arg \min_{(z_{1}, z_{0}) \in \mathbb{R}^{2}} \sum_{j=0,1} \left[ -k z_{j} + \sum_{i=1}^{n} W_{j}(D_{i,j}, P(X_{ij})) l_{\delta}(J_{i,j}, z_{j}) \right]. \quad (A.1.22) \]

Step 4:
Denote
\[ Q_{j,n}(z_{j}, k) = - \sum_{i=1}^{n} W_{j}(D_{i}, \hat{P}(X_{i})) k \tau_{n} z_{j} + \sum_{i=1}^{n} W_{j}(D_{i}, \hat{P}(X_{i})) l_{\delta}(\alpha_{j,n}(U_{i,j} - \beta_{j,n}), z_{j}) \]
and
\[ Q_{j,\omega}(z_j, k) = \left( -k z_j + \sum_{i=1}^{\infty} W_j(D_{i,j}, P(X_{i,j})) l_\delta(J_{i,j}, z_j) \right). \]

Then I have
\[ (\hat{Z}_{1,n}(k_l), \hat{Z}_{0,n}(k_l))_{l=1}^{L} = \arg\min_{(z_{1,\omega}, z_{0,\omega})_{l=1}^{L} \in \mathbb{R}^{2L}} \sum_{l=1}^{L} \sum_{j=0,1} Q_{j,n}(z_{j,l}, k_l). \]

When \( k_l \) satisfies Assumption 9 for \( l = 1, 2, \ldots, L \), by repeating Step 1–3, I can establish that
\[ \sum_{l=1}^{L} \sum_{j=0,1} Q_{j,n}(z_{j,l}, k_l) \sim \sum_{l=1}^{L} \sum_{j=0,1} Q_{j,\omega}(z_{j,l}, k_l). \]

By the same Convexity Lemma used in Step 3, I have
\[ (\hat{Z}_{1,n}(k_l), \hat{Z}_{0,n}(k_l))_{l=1}^{L} \sim (Z_{1,\omega}(k_l), Z_{0,\omega}(k_l))_{l=1}^{L} \]
\[ := \arg\min_{(z_{1,\omega}, z_{0,\omega})_{l=1}^{L+1} \in \mathbb{R}^{2(L+1)}} \sum_{l=1}^{L} \sum_{j=0,1} \left[ -k_l z_{j,l} + \sum_{i=1}^{n} W_j(D_{i,j}, P(X_{i,j})) l_\delta(J_{i,j}, z_{j,l}) \right]. \]

Step 5:
I aim to prove the result by applying Theorem 13.1 of Billingsley (1999) with \( T_p = [\kappa_1, \kappa_2]/(L_1 \cup L_0) \) because as mentioned above, all the discontinuities of the \( Z_{j,\omega}(k) \) occurs in \( L_j \). In fact, with \( (\kappa_1, \kappa_2) \notin L_1 \cup L_0 \), I only need to show \( (\hat{Z}_{1,n}(k_l), \hat{Z}_{0,n}(k_l)) \) indexed by \( k \in [\kappa_1, \kappa_2] \) is tight. Then based on Theorem 13.3 of Billingsley (1999), it suffices to show that \( T_p \)'s complement in \([\kappa_1, \kappa_2]\) is at most countable, that for \( j = 0,1 \) and every \( \varepsilon \),
\[ \lim_{\delta \to 0} \left[ P \left( |Z_{j,\omega}(\kappa_2) - Z_{j,\omega}(\kappa_2 - \delta)| \geq \varepsilon \right) + P \left( |Z_{j,\omega}(\kappa_1) - Z_{j,\omega}(\kappa_1 + \delta)| \geq \varepsilon \right) \right] = 0, \]
(A.1.23)
and that, for \( j = 0, 1 \), any positive \( \varepsilon \), and any \( \eta \), there exists constants \( \delta \) and \( n_0 \) such that

\[
P \left( |\omega_j^n(\delta)| \geq \varepsilon \right) \leq \eta \quad (A.1.24)
\]

in which

\[
\omega_j^n(\delta) := \sup_{k_1 \leq k_2 \leq k_3, k_3 - k_1 \leq 3} \left\{|\hat{Z}_{j,n}(k_2) - \hat{Z}_{j,n}(k_1)| \land |\hat{Z}_{j,n}(k_3) - \hat{Z}_{j,n}(k_2)|\right\}.
\]

(A.1.23) holds by Assumption 9. For (A.1.24), I focus on the case for \( j = 1 \). The case for \( j = 0 \) can be handled similarly. First, by convention, I define \( Z_{1,\infty}(k) \) as the left limiting of the sample path, that is, \( Z_{1,\infty}(k) = \lim_{k \downarrow k} Z_{1,\infty}(k') \). Then I notice that \( Z_{1,\infty}(k) \) is piece-wise constant and that the jumps only occur when

\[
k - \frac{1}{P(X_h)} = \sum_{i \neq h} \frac{T_i}{P(X_i)} 1\{J_i < J_h\}
\]

is finite. Hence by Theorem 13.3 of Billingsley (1999), the marginal processes \( \hat{Z}_{1,n}(k) \) and \( \hat{Z}_{2,n}(k) \) indexed by \( k \) in \( D[\kappa_1, \kappa_2] \) are tight and \( (\hat{Z}_{1,n}(k), \hat{Z}_{0,n}(k)) \) converges to \( (\hat{Z}_{1,\infty}(k), \hat{Z}_{0,\infty}(k)) \) under Skorohod metric.

A.1.8 Proof of Corollary 2.4.2

I only have to show the weak convergence of \( P(X \in \cdot | Y_1 = y) \) to \( \sum_t P(x_t) P_t \mathbb{1}\{x_t \in \cdot\} \), that is, for any \( F \in \text{Supp}(X) \) with \( \partial F \cap \{x_1, x_2, \ldots, x_T\} = \emptyset \),

\[
\lim_{y \to a_1(0)} P(X \in F| Y_1 = y) = \sum_{t=1}^T P_t \mathbb{1}\{x_t \in F\}.
\]

I first claim that for an arbitrarily small constant
\( \gamma \), there exist a small constant \( \eta \), such that for any \( t = 1, \cdots, T \), if \( |y - q_1(0)| < \eta \), \( S_{y,t} = \{ x : |x - x_t| \leq \gamma \} \).

Suppose not, since \( T \) is finite, as \( y \downarrow q_1(0) \), there exists a \( t \) and a sequence \( x_{y,t} \in S_{y,t} \), such that \( |x_{y,t} - x_t| > \gamma_0 \). Also because \( x_{y,t} \in S_y \), there exists a corresponding \( \varepsilon_{y,t} \) such that \( g(x_{y,t}, \varepsilon_{y,t}) \leq y \). Since \( \text{Supp}(X) \times [0,1] \) is compact, there is a convergent subsequence \( \{ x'_{y,t}, \varepsilon'_{y,t} \} \) of \( \{ x_{y,t}, \varepsilon_{y,t} \} \) with limiting point \( (x', \varepsilon') \). Since \( g(x'_{y,t}, \varepsilon'_{y,t}) \leq y' \) and \( g \) is lower semi-continuous, as \( y' \to q_1(0) \), \( g(x', \varepsilon') \leq \liminf_{y' \to q_1(0)} g(x'_{y,t}, \varepsilon'_{y,t}) = q_1(0) \). So \( g(x', \varepsilon') = q_1(0) \). This means \( x' \in S_0 \). But \( |x' - x_t| \geq \gamma_0 \). In addition, \( S_{y,t} \) is monotone decreasing in \( y \) by construction so \( \{ x'_{y,t} \} \subset S_{y_0,t} \). This implies \( d(x', S_{y_0,t}) = 0 \) for some \( t' \neq t \). This contradicts with the construction of \( S_{y_0,t} \).

Let \( \delta_0 = \min_{(x_t, x_{t'}) \in S_0 \times S_0} \| x_t - x_{t'} \| \) and \( B(x, d) \) be a ball with radius \( d \) and center \( x \). Then when \( y \) is small enough, \( S_{y,t} = S_y \cap B(x, \delta_0/2) \), which is defined independent of the initial partition \( \{ S_{y_0,t} \}_{t=1}^T \). This implies \( P_t \) is well defined independent of \( S_{y_0,t} \). Furthermore, for any \( F \) such that \( \partial F \cap \{ x_1, x_2, \cdots, x_T \} = \emptyset \), either \( d(x_t, F) > 0 \) or \( d(x_t, F^c) > 0 \) for all \( t = 1, 2, \cdots, T \). If \( d(x_t, F^c) > 0 \), \( s_{y,t} \subset F \) whenever \( y \) is small enough. If \( d(x_t, F) > 0 \), \( S_{y,t} \cap F = \emptyset \) whenever \( y \) is small enough. Therefore, for some arbitrarily small \( \gamma \), there always exists a \( y \) small enough such that

\[
|P(X \in F|Y_1 = y) - \sum_t P_t 1\{x_t \in F\}| \\
= \sum_{t=1}^T \left[ \frac{\mathbb{E}1\{X \in S_{y,t} \cap F\} \frac{\partial \lambda(X,Y)}{\partial y}}{\mathbb{E}1\{X \in S_y\} \frac{\partial \lambda(X,Y)}{\partial y}} - P_t 1\{x_t \in F\} \right] \\
\leq \sum_{t=1}^T |P_{y,t} - P_t 1\{x_t \in F\}| \\
\leq M \gamma
\]

This implies that \( P(X \in .|Y_1 = y) \) weakly converges to \( \sum_t P_t 1\{x_t \in .\} \).
A.1.9  Proof of Corollary 2.4.4

In the proof of corollary 2.4.2, I have shown that for any $\gamma > 0$, $S^d_{y,r} \in \mathcal{B}(x_r, \gamma)$. I next want to show it is also true for $S^c_{y,t}$, that is, $S^c_{y,t} \in (S_{0,t})^\gamma$.

Suppose not, then there exists $\gamma_0 > 0$ and a sequence $x_{y,t} \in S^c_{y,t}$ such that $d(x_{y,t}, S_{0,t}) > \gamma_0$. $x_{y,t} \in S^c_{y,t}$ implies that there exists a corresponding sequence $\{e_{y,t}\}$ such that $g(x_{y,t}, e_{y,t}) \leq y$. Then there exists a convergent subsequence $(x'_{y,t}, e'_{y,t})$ with limit $(x', e')$ such that $g(x', e') \leq \liminf_{y \to q_1(0)} g(x_{y,t}, e_{y,t}) \leq q_1(0)$. This implies $x' \in S_0$. But $d(x', S_{0,t}) > \gamma_0$, so $x' \in S_{0,t'}$ for $t' \neq t$ or $x' = x_r$, for some $r = 1, 2, \ldots, R^d$. But $S^c_{y,t}$ is decreasing so I have $d(x', S^c_{y_0,t}) = 0$. This contradicts with the way I construct $\{S^c_{y_0,t}\}_{t=1}^{T^c}$ and $\{S^d_{y_0,t}\}_{r=1}^{R^d}$.

The above claim implies that whenever $y$ is small enough, $S^d_{y,r} = \mathcal{B}(x_r, \delta_0/2) \cap S_y$ and $S^c_{y,t} = (S_{0,t})^{\delta_0/2} \cap S_y$. Then $\{S^d_{y,t}\}_{r=1}^{R^d}$ and $\{S^c_{y,t}\}_{t=1}^{T^c}$ are defined independent of $\{S^c_{y_0,t}\}_{t=1}^{T^c}$ and $\{S^d_{y_0,r}\}_{r=1}^{R^d}$ and they are disjoint. This implies $P^d_{y,t}$ and $P^c_{y,t}$ are well defined independent of $\{S^c_{y_0,t}\}_{t=1}^{T^c}$ and $\{S^d_{y_0,r}\}_{r=1}^{R^d}$. Furthermore, $S_{0,t}$ is compact because for a convergent sequence $\{x_n\}_{n=1}^{\infty}$ with limit $x$, there exists a corresponding sequence $\{\varepsilon_n\}_{n=1}^{\infty} \subset [0, 1]$ such that it has a convergent subsequence $\{\varepsilon'_{n'}\}$ with limit $\varepsilon$. Then $g(x, \varepsilon) \leq \liminf_{\varepsilon \to q_1(0)} g(x_{n'}, \varepsilon'_{n'}) \leq q_1(0)$, which implies $x \in S_0$. Since all $S_{0,t'}$, $t' = 1, 2, \ldots, T^c$ are separate, it implies $x \in S_{0,t}$. Therefore, $F \cap S^c_{y,t} \to F \cap S_{0,t}$.

The potential discontinuity $\mathcal{S}$ of the limiting distribution is $\{x_r\}_{r=1}^{R^d} \cup \left(\mathcal{S}_X \cap (\cup_{r=1}^{R^d}(\partial S_{0,r}))\right)$ where $\mathcal{S}_X$ is the discontinuity of $X$. Let $\mathcal{F}$ be a collection of all open and relatively compact set such that $\partial F \cap \mathcal{S} = \emptyset$. Then, in order to show the weak convergence, it suffices to show that

$$\lim_{y \to q_1(0)} P(X \in F|Y_1 = y) = \sum_r \mathbb{1}\{X_r \in F\} P^d_r + \sum_{t=1}^{T} P^c_t \int_{S_{0,t} \cap F} \int_{S_{0,t}} \frac{\sigma_t(x)^{1/\xi}\sigma_t(x)^{1/\xi}dF_X(x)}{\sigma_t(x)^{1/\xi}\sigma_t(x)^{1/\xi}dF_X(x)},$$

194
for all $F$ with $\partial F \cap S = \emptyset$.

Notice that $\frac{f_{t}(y-q_{t}(0)\mid X)}{f_{t}(y-q_{t}(0))} \rightarrow \sigma_{t}(X)^{-1/\xi_{t}}$ locally uniformly and $F \cap S_{y,t} \rightarrow F \cap S_{0,t}$.

Then, by the dominated convergence theorem, as $y \rightarrow q_{1}(0)$, I have

$$
\frac{\mathbb{E}\left\{ X \in F \cap S_{y,t} \right\} \frac{\delta (X,y)}{\delta y}}{\mathbb{E}\left\{ X \in S_{y,t} \right\} \frac{\delta (X,y)}{\delta y}} = \frac{\mathbb{E}\left\{ X \in F \mid S_{y,t} \right\} \frac{f_{t}(y-q_{t}(0)\mid X)}{f_{t}(y-q_{t}(0))} \delta (X,y)}{\mathbb{E}\left\{ X \in S_{y,t} \right\} \frac{f_{t}(y-q_{t}(0))}{f_{t}(y-q_{t}(0))} \delta (X,y)} \\
\rightarrow \frac{\mathbb{E}\left\{ X \in F \cap S_{0,t} \right\} \sigma_{t}(X)^{-1/\xi_{t}}}{\mathbb{E}\left\{ X \in S_{0,t} \right\} \sigma_{t}(X)^{-1/\xi_{t}}}.
$$

Therefore, for any fixed $F$ such that $\partial F \cap S = \emptyset$, as $y \rightarrow q_{1}(0)$,

$$
P(X \in F \mid Y = y)
= \frac{\mathbb{E}\left\{ X \in F \right\} \frac{\delta (X,y)}{\delta y}}{\mathbb{E}\left\{ X \in S_{y,t} \right\} \frac{\delta (X,y)}{\delta y}}
= \sum_{r=1}^{R} \frac{\mathbb{E}\left\{ X \in F \cap S_{y,r} \right\} \frac{\delta (X,y)}{\delta y}}{\mathbb{E}\left\{ X \in S_{y,r} \right\} \frac{\delta (X,y)}{\delta y}} + \sum_{t=1}^{T} \frac{\mathbb{E}\left\{ X \in F \cap S_{y,t} \right\} \frac{\delta (X,y)}{\delta y}}{\mathbb{E}\left\{ X \in S_{y,t} \right\} \frac{\delta (X,y)}{\delta y}}
= \sum_{r=1}^{R} \frac{P_{g,r}^{d} \mathbb{E}\left\{ X \in F \cap S_{y,r} \right\} \frac{\delta (X,y)}{\delta y}}{\mathbb{E}\left\{ X \in S_{y,r} \right\} \frac{\delta (X,y)}{\delta y}} + \sum_{t=1}^{T} \frac{P_{g,t}^{c} \mathbb{E}\left\{ X \in F \cap S_{y,t} \right\} \frac{\delta (X,y)}{\delta y}}{\mathbb{E}\left\{ X \in S_{y,t} \right\} \frac{\delta (X,y)}{\delta y}}
\rightarrow \sum_{r=1}^{R} P_{x,r}^{d} \chi_{x \in F} + \sum_{t=1}^{T} P_{t}^{c} \frac{\mathbb{E}\left\{ X \in F \cap S_{0,t} \right\} \sigma_{t}(X)^{-1/\xi_{t}}}{\mathbb{E}\left\{ X \in S_{0,t} \right\} \sigma_{t}(X)^{-1/\xi_{t}}}.
$$

This concludes the proof.
Similarly, results with Theorem 2.4.1, I obtain that

\[ \frac{\hat{\alpha}_n}{\alpha_{1,n}} \sim \frac{\sqrt{k_\rho}}{\max(\hat{Z}_{1,n}(mk_\rho) - \hat{Z}_{1,n}(k_\rho), \frac{\alpha_{1,n}}{\alpha_{0,n}}(\hat{Z}_{0,n}(mk_\rho) - \hat{Z}_{0,n}(k_\rho)))} \]

Similarly, \( \frac{\hat{\alpha}_n}{\alpha_{1,n}} \sim \frac{\sqrt{k_\rho \tilde{\rho}}}{\max(\hat{Z}_{1,n}(mk_\rho) - \hat{Z}_{1,n}(k_\rho), \tilde{\rho}(\hat{Z}_{0,n}(mk_\rho) - \hat{Z}_{0,n}(k_\rho)))} \). By combining the above results with Theorem 2.4.1, I obtain that

\[ \hat{Z}_n(k) = \hat{\alpha}_n(\hat{q}(\tau_n) - q(\tau_n)) = \frac{\hat{\alpha}_n}{\alpha_{1,n}} \hat{Z}_{1,n}^c(k) - \frac{\hat{\alpha}_n}{\alpha_{0,n}} \hat{Z}_{0,n}^c(k) \quad \sim Z_{\infty}^c(k). \]

The limiting distribution is non-degenerate even when \( \rho = 0 \) or \( \infty \).

**A.1.11 Proof of Proposition 2.4.5**

\[
\hat{\alpha}_n \left( \sum_{l=1}^{L} \hat{r}_l \hat{q}(\tau_{n,l}) - \sum_{l=1}^{L} r_l q(\tau_{n,l}) \right) \\
= \hat{\alpha}_n \left( \sum_{l=1}^{L} (\hat{r}_l - r_l) q(\tau_{n,l}) \right) + \hat{\alpha}_n \left( \sum_{l=1}^{L} \hat{r}_l (\hat{q}(\tau_{n,l}) - q(\tau_{n,l})) \right) \\
= \hat{\alpha}_n \left( \sum_{l=1}^{L} (\hat{r}_l - r_l) q(\tau_{n,l}) - q(0) \right) + \hat{\alpha}_n \left( \sum_{l=1}^{L} \hat{r}_l (\hat{q}(\tau_{n,l}) - q(\tau_{n,l})) \right).
\]

Since \( \alpha_{j,n}(q_j(\tau_n) - q_j(0)) \to \eta_j(k) \), \( \frac{\hat{\alpha}_n}{\alpha_{j,n}} = O_p(1) \) for \( j = 0, 1 \), and \( \hat{\gamma}_l \to \gamma_l \), the first term is \( O_p(1) \). The second term converges to \( \sum_{l=1}^{L} \gamma_l Z_{\infty}^c(k_\rho) \). This concludes the proof.

**A.1.12 Proof of Theorem 2.5.1**

The proof follows the five steps in the proof of Theorem 2.4.1 which I will not repeat. The key ingredient, Lemma A.1.2, is replaced by the following Lemma.
Lemma A.1.3. Let $P_{n,i} = \sum_{i=1}^{n} 1\{I_i = i\}$. Under the conditions of Theorem 2.5.1, for $j = 0, 1$,

1. $\frac{1}{n} \sum_{i=1}^{n} P_{n,i} W_j(D_i, P(X_i)) \to 1$ a.s.
2. For $\hat{N}_j^* := \sum_{i=1}^{n} P_{n,i} 1\{\alpha_{j,n}(U_{i,j} - \beta_{j,n}), D_i, X_i\}$,

$$\hat{N}_j^* \rightsquigarrow N_j^* := \sum_{i=1}^{\infty} \Gamma_{i,j} 1\{J_{i,j}, D_{i,j}, X_{i,j}\}.$$

3. Let

$$g_1(u, x) = \frac{1}{P(x)} l_6(u, x, z_1), \quad g_0(u, x) = \frac{1}{1 - P(x)} l_6(u, x, z_0),$$

and

$$\Psi_j,n = \sum_{i=1}^{n} (j D_i + (1 - j)(1 - D_i)) P_{n,i} g_j(\alpha_{j,n}(U_{i,j} - \beta_{j,n}), X_i).$$

Then for a pair of constants $(t_1, t_0)$,

$$\mathbb{E} \exp(\tilde{t}_1 \Psi_{1,n} + \tilde{t}_0 \Psi_{0,n}) \to \mathbb{E} \exp(\tilde{t} \int_{E_1} t_1 d\tilde{g}_1 dN_1^*) \mathbb{E} \exp(\tilde{t} \int_{E_0} t_0 (1 - d) g_0 dN_0^*),$$

in which $N_j$ is defined in (2).

4. The distances between the two closest discontinuities of the marginal sample paths of the two-dimensional stochastic process $(\hat{Z}_{1,n}^*(k), \hat{Z}_{0,n}^*(k))$ indexed by $k$ are both greater than 1.

A.1.13 Proof of Theorem 2.5.2

The proof is divided into three steps. For $j = 0, 1$, denote $Z_{j,n}^*(k) = \alpha_{j,n}(\tilde{q}_j^*(\tau_0) - q_j(0))^2$ where $\alpha_{j,n}$ is the infeasible convergence rate defined after Assumption 7. In the first step, I want to show that $(Z_{1,n}^*(k), Z_{0,n}^*(k))$ as a two-dimensional stochastic process indexed by $k$ in $\mathcal{D}^2([\kappa_1, \kappa_2])$ converges weakly to $(Z_{1,\infty}(k), Z_{0,\infty}(k))$ defined  

$^2$ It is different from $\hat{Z}_{n}^*(k) = \alpha_{n}(\tilde{q}_n^*(\tau_0) - q_n(\tau_0)). \tilde{q}_n^*(\tau_0)$ is defined before Theorem 2.5.2.
in Theorem 2.4.1 under Skorohod metric. In the second step, I want to show that 
\[ \hat{\alpha}^*_b(\hat{q}_b^*(\tau_b) - q(\tau_b)) \] as a two-dimensional stochastic process indexed by \( k \) in \( \mathcal{D}([\kappa_1, \kappa_2]) \) converges weakly to \( Z^c_\infty(k) \) defined in Theorem 2.4.2 under Skorohod metric. Last, I want to show that 
\[ \hat{\alpha}^*_b(\hat{q}_n(\tau_b) - q(\tau_b)) \] as a stochastic process indexed by \( k \) in \( \mathcal{D}([\kappa_1, \kappa_2]) \) converges weakly to 0 under uniform metric. Combining the results from the last two steps, I can establish the desired result that

\[ \hat{\alpha}^*_b(\hat{q}_b^*(\tau_b) - \hat{q}_n(\tau_b)) = \hat{\alpha}^*_b(\hat{q}_b^*(\tau_b) - q(\tau_b)) \rightsquigarrow Z^c_\infty(k). \]

Step 1.

\[
(\hat{Z}_{1,b}^*(k), \hat{Z}_{0,b}^*(k)) = \arg \min_{(z_1, z_2)} \left\{ -\sum_{j=0,1} -\sum_{i=1}^n P_{n,i} W_j(D_i, \hat{P}(X_i)) \tau_b z_j + \sum_{i=1}^n P_{n,i} W_j(D_i, \hat{P}(X_i)) l_b(\alpha_{j,b}(U_{i,j} - q(0)), z_j) \right\}.
\]

If the replacement is allowed, \( P_{n,i} = \sum_{i=1}^b \mathbb{1}\{I_l = i\} \), \( (I_{n,1}, I_{n,2}, \ldots, I_{n,b}) \) is a multinomial vector with parameter \( b \) and probabilities \( \left( \frac{1}{n}, \ldots, \frac{1}{n} \right) \). If replacement is not allowed, \( \{P_{n,i}\}^{n}_{i=1} \) has \( b \) 1’s and \( n-b \) 0’s and each combination of \( \{P_{n,i}\}^{n}_{i=1} \) has equal probability \( \frac{1}{C_n^b} \). The proof of this step follows the five steps in the proof of Theorem 2.4.1 which I will not repeat. The key ingredient, Lemma A.1.2, is replaced by the following Lemma.

**Lemma A.1.4.**

(1) \( \frac{1}{n} \sum_{i=1}^n P_{n,i} W_j(D_i, P(X_i)) \to 1 \) a.s.

(2) For \( \hat{N}_j^* := \sum_{i=1}^n P_{n,i} \mathbb{1}\{\alpha_{j,b}(U_{i,j} - \beta_{j,b}), D_i, X_i\} \),

\[ \hat{N}_j^* \rightsquigarrow N_j := \sum_{i=1}^\infty \mathbb{1}\{J_{i,j}, D_{i,j}, X_{i,j}\}. \]
(3) Let
\[ g_1(u, x) = \frac{1}{P(x)}l_\delta(u, x, z_1), \quad g_0(u, x) = \frac{1}{1 - P(x)}l_\delta(u, x, z_0), \]

and
\[ \Psi_{j,n} = \sum_{i=1}^{n}(jD_i + (1 - j)(1 - D_i))P_{n,i}g_j(\alpha_{j,b}(U_{i,j} - \beta_{j,b}), X_i). \]

Then for a pair of constants \((t_1, t_0)\),
\[ \mathbb{E}\exp(i\tilde{t}_1\Psi_{1,n} + i\tilde{t}_0\Psi_{0,n}) = \mathbb{E}\exp(\tilde{t}_1\int_{E_1} t_1 dg_1 dN_1)\mathbb{E}\exp(\tilde{t}_0\int_{E_0}(1 - d)g_0 dN_0), \]
in which \(N_j\) is defined in (2).

(4) The distances between the two closest discontinuities of the marginal sample paths of the two-dimensional stochastic process \((\hat{Z}_{1,n}^*(k), \hat{Z}_{0,n}^*(k))\) indexed by \(k\) are both greater than 1.

Step 2.
First, I note that
\[ \hat{\alpha}_b^*(\hat{q}_b^*(\tau_b) - q(0)) = \frac{\alpha_b^*}{\alpha_{1,b}}\hat{Z}_{1,b}^*(k) - \frac{\alpha_b^*}{\alpha_{0,b}}\hat{Z}_{0,b}^*(k), \]
\[ \alpha_{1,b} \max(\hat{q}_b^*(m\tau_{b,r}) - \hat{q}_1^*(\tau_{b,r}), \hat{q}_0^*(m\tau_{b,r}) - \hat{q}_0^*(\tau_{b,r})) \]
\[ \rightarrow \max(Z_{1,\infty}(mk_{r'}) - Z_{1,\infty}(k_{r'}), \hat{\rho}(Z_{0,\infty}(mk_{r'}) - Z_{0,\infty}(k_{r'}))), \]
and similarly,
\[ \alpha_{0,b} \max(\hat{q}_b^*(m\tau_{b,r}) - \hat{q}_1^*(\tau_{b,r}), \hat{q}_0^*(m\tau_{b,r}) - \hat{q}_0^*(\tau_{b,r})) \]
\[ \rightarrow \max(\frac{1}{\hat{\rho}}(Z_{1,\infty}(mk_{r'}) - Z_{1,\infty}(k_{r'})), Z_{0,\infty}(mk_{r'}) - Z_{0,\infty}(k_{r'})). \]

By step 1, I have
\[ (\hat{Z}_{1,b}^*(k), \hat{Z}_{0,b}^*(k)) \quad \rightarrow \quad (Z_{1,\infty}(k), Z_{0,\infty}(k)). \]
Therefore
\[
\hat{\alpha}_b^*(\hat{q}_b^*(\tau_b) - q(\tau_b)) \xrightarrow{\text{d}} \frac{\sqrt{k^p}(Z_{1,\infty}(k) - \hat{\rho}Z_{0,\infty}(k))}{\max(Z_{1,\infty}(mk^p) - Z_{1,\infty}(k^p), \hat{\rho}(Z_{0,\infty}(mk^p) - Z_{0,\infty}(k^p)))}.
\]

Last, I have that \(\alpha_{j,b}(q_j(\tau_b) - q_j(0)) \to \eta_j(k)\) uniformly in \(k \in [\kappa_1, \kappa_2]\). Combining this with the above result, I obtain that
\[
\hat{\alpha}_b^*(\hat{q}_b^*(\tau_b) - q(\tau_b)) \xrightarrow{\text{d}} \frac{\sqrt{k^p}(Z_{1,\infty}(k) - \hat{\rho}Z_{0,\infty}(k))}{\max(Z_{1,\infty}(mk^p) - Z_{1,\infty}(k^p), \hat{\rho}(Z_{0,\infty}(mk^p) - Z_{0,\infty}(k^p)))}.\]

This concludes step 2.

Step 3.
By construction, \(\tau_b n = \tau_n n_b^2 \to \infty\). By Theorem 2.3.1, \(\lambda_{j,n}(k)(\hat{q}_j(\tau_b) - q_j(\tau_b))\) as a stochastic process indexed by \(k\) is tight. I only need to show \(\frac{\hat{\alpha}_b^*}{\lambda_{j,n}(k)} \to 0\). To see this, I note that, by step 1, \(\hat{\alpha}_b^* = O_p(\min(\alpha_{1,b}, \alpha_{0,b}))\). Furthermore, since \(k \in [\kappa_1, \kappa_2]\), I have
\[
\frac{\hat{\alpha}_b^*}{\lambda_{j,n}(k)} \leq_p \frac{\alpha_{j,b}}{\lambda_{j,n}(k)} \leq_p \sqrt{\frac{b}{n\kappa_1}} = o(1).
\]

This concludes the proof.

A.1.14 Proof of Corollary 2.5.1

By Assumption 18 and Theorem 2.5.2, I have
\[
\hat{Z}_n^c(k)/S_n(k) \lawlim Z_{\infty}^c(k)/\sigma(k) \text{ in } \mathcal{D}[\kappa_1, \kappa_2].
\]

Let \(\rho\) be the Skorohod metric on \(\mathcal{D}([\kappa_1, \kappa_2])\). Since 0 is a constant function, the map \(\rho(s, 0) = \sup_{k \in [\kappa_1, \kappa_2]} |s|\) is continuous in \(s \in \mathcal{D}([\kappa_1, \kappa_2])\). Therefore,
\[
\sup_{k \in [\kappa_1, \kappa_2]} |\hat{Z}_n^c(k)/S_n(k)| \lawlim \sup_{k \in [\kappa_1, \kappa_2]} |Z_{\infty}^c(k)/\sigma(k)|.
\]
Next, I note that \( \sup_{k \in [\kappa_1, \kappa_2]} |Z_{\infty}^c(k)/\sigma(k)| \) is continuously distributed by Lemma A.1.8. Thus,

\[
    \tilde{C}_{1-a} \xrightarrow{p} C_{1-a}
\]

in which \( \tilde{C}_{1-a} \) and \( C_{1-a} \) are the \((1 - a)\)-th quantiles of

\[
    \sup_{k \in [\kappa_1, \kappa_2]} |\hat{Z}^c_{\tau_n}(k)/S_n(k)| \quad \text{and} \quad \sup_{k \in [\kappa_1, \kappa_2]} |Z_{\infty}^c(k)/\sigma(k)|,
\]

respectively.

This implies that the \((1 - a)\)-th uniform confidence band is consistent, that is,

\[
    \lim_{n \to \infty} P \left( q\left( \frac{k}{n} \right) \in \left[ \hat{q}\left( \frac{k}{n} \right) - S_n(k)\tilde{C}_{1-a}/\hat{\alpha}_n, \hat{q}\left( \frac{k}{n} \right) + S_n(k)\tilde{C}_{1-a}/\hat{\alpha}_n \right] : k \in [\kappa_1, \kappa_2] \right) = 1 - \alpha.
\]

A.1.15 Proof of Theorem 2.5.3

If \( \{\tau_n\}_{n \geq 1} \in \Gamma_{ex} \) and \( \tau_n \leq \tau_{n,1} \) for \( n \) large enough,

\[
    \tilde{C}^h_a(\tau_n) = \hat{C}^{bn}_a(\tau_n).
\]

By Theorem 2.5.1,

\[
    P \left( q(\tau_n) \in \left( \hat{q}(\tau_n) - \tilde{C}^h_{1-\frac{a}{2}}(\tau_n), \hat{q}(\tau_n) - \tilde{C}^{bn}_{\frac{a}{2}}(\tau_n) \right) \right) = 1 - a.
\]

If \( \{\tau_n\}_{n \geq 1} \in \Gamma_{ex} \) and for \( n \) large enough, \( \tau_n > \tau_{n,1} \),

\[
    \tilde{C}^h_a(\tau_n) = \check{C}^{ds}_a(\tau_n)
\]

and thus

\[
    P \left( q(\tau_n) \in \left( \hat{q}(\tau_n) - \tilde{C}^h_{1-\frac{a}{2}}(\tau_n), \hat{q}(\tau_n) - \tilde{C}^{bn}_{\frac{a}{2}}(\tau_n) \right) \right) 
\geq P \left( q(\tau_n) \in \left( \hat{q}(\tau_n) - \check{C}^{bn}_{1-\frac{a}{2}}(\tau_n), \hat{q}(\tau_n) - \check{C}^{bn}_{\frac{a}{2}}(\tau_n) \right) \right) = 1 - a.
\]

These two situations exhaust all sequences in \( \Gamma_{ex} \).
If \( \{\tau_n\}_{n \geq 1} \in \Gamma_{\text{int}} \), for \( n \) large enough, I have \( \tau_n \geq \tau_{n,1} \). This implies that

\[
P \left( q(\tau_n) \in \left( \hat{q}(\tau_n) - \hat{C}_{1-\frac{a}{2}}^h(\tau_n), \hat{q}(\tau_n) - \hat{C}_{\frac{a}{2}}^h(\tau_n) \right) \right)
\]

\[
\geq P \left( q(\tau_n) \in \left( \hat{q}(\tau_n) - \hat{C}_{1-\frac{a}{2}}^m(\tau_n), \hat{q}(\tau_n) - \hat{C}_{\frac{a}{2}}^m(\tau_n) \right) \right) = 1 - a,
\]

where the last equality is by Theorem 2.3.3.

If \( \{\tau_n\}_{n \geq 1} \in \Gamma_{\text{reg}} \), for \( n \) large enough, I have \( \tau_n \geq \tau_{n,2} \). This implies that

\[
\hat{C}_a^h(\tau_n) = \hat{C}_a^m(\tau_n),
\]

and thus by the assumption in the theorem,

\[
P \left( q(\tau_n) \in \left( \hat{q}(\tau_n) - \hat{C}_{1-\frac{a}{2}}^h(\tau_n), \hat{q}(\tau_n) - \hat{C}_{\frac{a}{2}}^h(\tau_n) \right) \right) = 1 - a.
\]

**A.1.16 Proof of Proposition 2.5.2**

It suffices to show that \( \hat{\alpha}_n(\hat{q}(0) - q(0)) \rightsquigarrow \sum_{l=1}^L \gamma_l Z^c_{\alpha}(k_l) \). Then Proposition 2.5.1 shows that \( \hat{C}_a \) is consistent for the \( \alpha \)-th quantile of \( \sum_{l=1}^L \gamma_l Z^c_{\alpha}(k_l) \).

First, by Theorem 2.3.4, \( \hat{\xi}_j \xrightarrow{p} \xi_j \) for \( j = 0, 1 \). This implies that \( (\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3) \xrightarrow{p} (\gamma_1, \gamma_2, \gamma_3) \) where \( (\gamma_1, \gamma_2, \gamma_3) \) is the unique solution to the following system of equations:

\[
\begin{align*}
3 \sum_{l=1}^3 r_l &= 1, & 3 \sum_{l=1}^3 r_l k_l^{-\xi_l} &= 0, & 3 \sum_{l=1}^3 r_l k_l^{-\xi_l} &= 0.
\end{align*}
\]

(A.1.25)

In addition,

\[
\hat{\alpha}_n(\hat{q}(0) - q(0)) = \hat{\alpha}_n \left\{ \sum_{l=1}^3 r_l [\hat{q}(\tau_{n,l}) - r_l q(\tau_{n,l})] \right\} + \hat{\alpha}_n \left\{ \sum_{l=1}^3 r_l [q(\tau_{n,l}) - q(0)] \right\}.
\]

Since \( (\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3) \xrightarrow{p} (\gamma_1, \gamma_2, \gamma_3) \), by Proposition 2.4.5, the first term converges weakly to \( \sum_{l=1}^L \gamma_l Z^c_{\alpha}(k_l) \). For the second term, since \( \alpha_{j,n}(q_j(\tau_{n,l}) - q_j(0)) \to \eta_j(k_l) = k_l^{-\xi_j} \)
and \( \hat{\alpha}_{n} = O_{p}(1) \), by (A.1.25), I have

\[
\hat{\alpha}_{n} \left\{ \sum_{t=1}^{3} r_{t}[q(\tau_{n,t}) - q(0)] \right\} = \left( \frac{\hat{\alpha}_{n}}{\hat{\alpha}_{1,n}} + \frac{\hat{\alpha}_{n}}{\hat{\alpha}_{0,n}} \right) o(1) = o_{p}(1).
\]

This concludes the proof.

\textbf{A.1.17 Proof of Lemma A.1.2}

(1) is trivial.

For (2), it is known that a Poisson random measure (PRM) with the Lebesgue mean measure can be written as

\[
\int F \mu_{j}(a,b) dR \times \{d\} \times F = \int_{F} \{dP(x) + (1 - d)(1 - P(x))\} P_{j}^{+}(dx) Y_{j} = q_{j}(0)) (h_{j}(b) - h_{j}(a)).
\]

I focus on \( j = 1 \). Since \( P_{j}^{+}(X \in |Y_{1} = q_{1}(0)) \) is a bounded measure, its discontinuities are at most countable. So there exists \( \mathcal{F}_{1} \), a basis of relatively compact open sets of \( \mathbb{R}^{d} \) such that \( \mathcal{F}_{1} \) is closed under finite unions and intersections and for any \( F \in \mathcal{F}_{1} \), \( P_{1}^{+}(X \in \partial F|Y_{1} = q_{1}(0)) = 0 \). Then by Lemma 9.3 and 9.4 in Chernozhukov (2005a), I only have to verify that, for any \( F \in \mathcal{F}_{1} \) and any interval \((a,b), \mathbb{E}N_{1}((a,b) \times \{d\} \times F) \to \mu_{1}((a,b) \times \{d\} \times F). \]

Notice that \( l/\alpha_{1,n} + \beta_{1,n} \downarrow F_{u_{1}}^{-1}(0) = 0 \) or \(-\infty\) for any \( l \in (-\infty, +\infty) \) for type 1 tails, any \( l \in (-\infty, 0) \) for type 2 tails, and \( l \in [0, +\infty) \) for type 3 tails. Let \( S_{n} = (q_{1}(0) + \beta_{1,n} + a/\alpha_{1,n}, q_{1}(0) + \beta_{1,n} + b/\alpha_{1,n}) \), by
the continuous mapping theorem, I obtain that

\[ \mathbb{E} \tilde{N}_1((a, b) \times \{d\} \times F) \]

\[ = P(D = d, X \in F|\alpha_{1,n}(U_1 - \beta_{1,n}) \in (a, b)) n P(\alpha_{1,n}(U_1 - \beta_{1,n}) \in (a, b)) \]

\[ = (1 + o(1)) \frac{\int_{S_n} P(D = d, X \in F|Y_1 = y) f_1(y) dy}{\int_{S_n} f_1(y) dy} (h_1(b) - h_1(a)) \]

\[ = (1 + o(1)) \frac{\int_{S_n \times F} (1 - d)(1 - P(x))) P(dx|Y_1 = y) f_1(y) dy}{\int_{S_n} f_1(y) dy} (h_1(b) - h_1(a)) \]

\[ \rightarrow \int_F (1 - d)(1 - P(x))) P_1^+(dx|Y_1 = q_1(0))(h_1(b) - h_1(a)). \]

This is the desired result for the marginal convergence.

For (3), let \((U'_{i,j}, X'_{i,j})_{j=0,1}\) be an i.i.d. sequence such that \((U'_{i,1}, X'_{i,1}) \perp (U'_{i,0}, X'_{i,0})\) and that \((U'_{i,j}, X'_{i,j})\) is distributed as \((U_{i,j}, X_i)|D_i = j\). Let \(p = P(D_i = 1)\). Then

\[ \mathbb{E} \exp(\tilde{t}_1 \Psi_{1,n} + \tilde{t}_0 \Psi_{0,n}) 1\{D_1 = 1, \cdots, D_s = 1, D_{s+1} = 0, \cdots, D_n = 0\} \]

\[ = \mathbb{E} \exp \left( \tilde{t}_1 \left( \sum_{i=1}^{s} g_1(\alpha_{1,n}(U_{i,1} - \beta_{1,n}), X_i) \right) + \tilde{t}_0 \left( \sum_{i=s+1}^{n} g_0(\alpha_{0,n}(U_{i,0} - \beta_{0,n}), X_i) \right) \right) \]

\[ \times 1\{D_1 = 1\} \cdots 1\{D_s = 0\} \cdot \prod_{i=s+1}^{n} 1\{D_i = 0\} \]

\[ = p^s(1 - p)^{n-s} \mathbb{E} \exp \left( \tilde{t}_1 \left( \sum_{i=1}^{s} g_1(\alpha_{1,n}(U'_{i,1} - \beta_{1,n}), X'_{i,1}) \right) \right) \]

\[ \times \mathbb{E} \exp \left( \tilde{t}_0 \left( \sum_{i=s+1}^{n} g_0(\alpha_{0,n}(U'_{i,0} - \beta_{0,n}), X'_{i,0}) \right) \right). \]
Therefore, by symmetry,
\[
\mathbb{E} \exp(\tilde{t}_1 \Psi_{1,n} + \tilde{t}_0 \Psi_{0,n})
\]
\[
= \sum_{s=0}^{n} C_n^s p^s (1 - p)^{n-s} \mathbb{E} \exp \left( \tilde{t}_1 \left( \sum_{i=1}^{s} g_1(\alpha_{1,n}(U'_{i,1} - \beta_{1,n}), X'_{i,1}) \right) \right)
\]
\[
\times \mathbb{E} \exp \left( \tilde{t}_0 \left( \sum_{i=s+1}^{n} g_0(\alpha_{0,n}(U'_{i,0} - \beta_{0,n}), X'_{i,0}) \right) \right).
\]

Define $E'_j$ for $j = 0, 1$ as follows:

- for type 1 tails ($\xi_j = 0$): $E'_j = [-\infty, +\infty) \times \text{Supp}(\mathcal{X})$,
- for type 2 tails ($\xi_j > 0$): $E'_j = [-\infty, 0) \times \text{Supp}(\mathcal{X})$,
- for type 3 tails ($\xi_j < 0$): $E'_j = [0, +\infty) \times \text{Supp}(\mathcal{X})$.

Let $N'_j$ be $\text{PRM}(\mu'_j)$ on $E'_j$ with
\[
\mu'_j([a, b] \times F) = \int_F (jP(x) + (1 - j)(1 - P(x))) P_j^+(dx|Y_j = q_j(0))(h_j(b) - h_j(a))
\]
and
\[
\tilde{N}'_j(.) := \sum_{i=1}^{js+(n-s)(1-j)} \mathbb{1} \{(\alpha_{j,n}(U'_{i,j} - \beta_{j,n}), X'_{i,j}) \in .\}.
\]
Let \( r_n = \sqrt{2n \log(\log(n))} \), \( S_n = \{ s \in \mathbb{Z}, |s - np| \leq r_n \} \). Then,

\[
\left| \mathbb{E} \exp(\tilde{i}t_1 \Psi_{1,n} + \tilde{i}t_0 \Psi_{0,n}) - \mathbb{E} \exp(\tilde{i} \int_{E_1} t_1 g_1 dN'_1) \mathbb{E} \exp(\tilde{i} \int_{E_0} t_0 g_0 dN'_0) \right|
\]

\[
\leq \sum_{s \in S_n} C_n^s p^s (1 - p)^{n-s} \left| \mathbb{E} \exp \left( \tilde{i}t_1 \int_{E_1} g_1 d\tilde{N}'_1 \right) \mathbb{E} \exp \left( \tilde{i}t_0 \int_{E_0} g_0 d\tilde{N}'_0 \right) \right|
\]

\[
- \mathbb{E} \exp \left( \tilde{i} \int_{E_1} t_1 g_1 dN'_1 \right) \mathbb{E} \exp \left( \tilde{i} \int_{E_0} t_0 g_0 dN'_0 \right)
\]

\[
+ \sum_{s \in S'_n} C_n^s p^s (1 - p)^{n-s} \left| \mathbb{E} \exp \left( \tilde{i}t_1 \int_{E'_1} g_1 d\tilde{N}'_1 \right) \mathbb{E} \exp \left( \tilde{i}t_0 \int_{E'_0} g_0 d\tilde{N}'_0 \right) \right|
\]

\[
- \mathbb{E} \exp \left( \tilde{i} \int_{E'_1} t_1 g_1 dN'_1 \right) \mathbb{E} \exp \left( \tilde{i} \int_{E'_0} t_0 g_0 dN'_0 \right)
\]

\[
\leq \sum_{s \in S_n} C_n^s p^s (1 - p)^{n-s} \left| \mathbb{E} \exp \left( \tilde{i}t_1 \int_{E_1} g_1 d\tilde{N}'_1 \right) \mathbb{E} \exp \left( \tilde{i}t_0 \int_{E_0} g_0 d\tilde{N}'_0 \right) \right|
\]

\[
- \mathbb{E} \exp \left( \tilde{i} \int_{E_1} t_1 g_1 dN'_1 \right) \mathbb{E} \exp \left( \tilde{i} \int_{E_0} t_0 g_0 dN'_0 \right) + \text{const} \times \left( \sum_{s \in S'_n} C_n^s p^s (1 - p)^{n-s} \right).
\]

(A.1.26)

By the law of iterated logarithm, \( \sum_{s \in S'_n} C_n^s p^s (1 - p)^{n-s} = o(1) \) as \( n \to \infty \). There-
fore, the second term is asymptotically negligible. For the first term, if \( s \geq [np] \),

\[
\left| \mathbb{E} \exp(i \int_{E_1} g_1 d \tilde{N}_1) - \mathbb{E} \exp(i \int_{E_1} g_1 d N'_1) \right|
\leq \left| \mathbb{E} \exp\left( i \int_{E_1} g_1 d \tilde{N}_1 \right) - \mathbb{E} \exp\left( i \int_{E_1} \sum_{i=1}^{[np]} g_1(\alpha_{1,n}(U'_{i,1} - \beta_{1,n}), X'_{i,1}) \right) \right|
\]

\[
+ \left| \mathbb{E} \exp\left( i \int_{E_1} \sum_{i=1}^{[np]} g_1(\alpha_{1,n}(U'_{i,1} - \beta_{1,n}), X'_{i,1}) \right) - \mathbb{E} \exp\left( i \int_{E_1} g_1 d N'_1 \right) \right|
\]

\[
\leq \left| \mathbb{E} \exp\left( i \int_{E_1} \sum_{i=1}^{[np]} g_1(\alpha_{1,n}(U'_{i,1} - \beta_{1,n}), X'_{i,1}) \right) \right|
\times \left| \exp\left( i \int_{E_1} \sum_{i=[np]}^{s} g_1(\alpha_{1,n}(U'_{i,1} - \beta_{1,n}), X'_{i,1}) \right) - 1 \right|
\]

\[
+ \left| \mathbb{E} \exp\left( i \int_{E_1} \sum_{i=1}^{[np]} g_1(\alpha_{1,n}(U'_{i,1} - \beta_{1,n}), X'_{i,1}) \right) - \mathbb{E} \exp\left( i \int_{E_1} g_1 d N'_1 \right) \right|
\]

\[
\leq \mathbb{E} \left( 2 - 2 \cos(t_1 \sum_{i=[np]+1}^{s} g_1(\alpha_{1,n}(U'_{i,1} - \beta_{1,n}), X'_{i,1})) \right)^{1/2}
\]

\[
+ \left| \mathbb{E} \exp\left( i \int_{E_1} \sum_{i=1}^{[np]} g_1(\alpha_{1,n}(U'_{i,1} - \beta_{1,n}), X'_{i,1}) \right) - \mathbb{E} \exp\left( i \int_{E_1} g_1 d N'_1 \right) \right|
\]

in which the last inequality is by the fact that \(|\exp(it) - 1|^2 \leq 2 - 2 \cos(t)|. \]
Similar to the proof in step 1,
\[
[np] P(\alpha_{1,n}(U_{i,1}' - \beta_{1,n}) \in [a, b], X_{i,1}' \in F)
\]
\[
= \frac{[np]}{p} P(\alpha_{1,n}(U_{i,1}' - \beta_{1,n}) \in [a, b], X_{i,1}' \in F, D_i = 1)
\]
\[
= \frac{[np]}{p} \int_a^b \int_F P(x) P(dx|\alpha_{1,n}(U_{i,1}' - \beta_{1,n}) = u) dP(\alpha_{1,n}(U_{i,1}' - \beta_{1,n}) \leq u)
\]
\[
= \mu_\gamma([a, b] \times F).
\]

Then by the continuous mapping theorem and the fact that \( g_1(u, x) \in C_k(E)' \), I have
\[
\sum_{i=1}^{[np]} g_1(\alpha_{1,n}(U_{i,1}' - \beta_{1,n}), X_{i,1}') \rightsquigarrow \int_{E_1} g_1 dN_1.'
\]

Similarly, because \( \frac{m_\alpha}{n} \to 0 \), I have that
\[
\sum_{i=[np]+1}^{s} |g_1(\alpha_{1,n}(U_{i,1}' - \beta_{1,n}), X_{i,1}')| \leq \sum_{i=[np]+1}^{[np+p_n]+1} |g_1(\alpha_{1,n}(U_{i,1}' - \beta_{1,n}), X_{i,1}')| = o_p(1).
\]
Therefore, for the first term on the RHS of (A.1.27), I have

\[
\sup_{s \in S_n, s \geq [np]} \left( 2 - 2 \cos(t_1 \sum_{i=[np]+1}^{[np+r_n]+1} g_1(\alpha_{1,n}(U'_{i,1} - \beta_{1,n}), X'_{i,1})) \right)
\]

\[
\leq 2 \left( 1 - \cos(|t_1| \sum_{i=[np]+1}^{[np+r_n]+1} |g_1(\alpha_{1,n}(U'_{i,1} - \beta_{1,n}), X'_{i,1})|) \right)
\]

\[
\times \left\{ \sum_{i=[np]+1}^{[np+r_n]+1} |g_1(\alpha_{1,n}(U'_{i,1} - \beta_{1,n}), X'_{i,1})| \leq \frac{\pi}{|t_1|} \right\}
\]

\[
+ 2\left\{ \sum_{i=[np]+1}^{[np+r_n]+1} |g_1(\alpha_{1,n}(U'_{i,1} - \beta_{1,n}), X'_{i,1})| \geq \frac{\pi}{|t_1|} \right\}
\]

\[= o_p(1).\]

Therefore, by the dominated convergence theorem, I have

\[
\sup_{s \in S_n, s \geq [np]} \mathbb{E} \left( 2 - 2 \cos(t_1 \sum_{i=[np]+1}^{[np]} g_1(\alpha_{1,n}(U'_{i,1} - \beta_{1,n}), X'_{i,1})) \right) \to 0.
\]

For the second term of (A.1.27), I have, by the dominated convergence theorem, that

\[
\left| \mathbb{E} \exp \left( \tilde{g}_1 \sum_{i=1}^{[np]} g_1(\alpha_{1,n}(U'_{i,1} - \beta_{1,n}), X'_{i,1}) \right) - \mathbb{E} \exp \left( \tilde{g}_1 \int_{E_1} g_1 dN_1' \right) \right| \to 0.
\]

Combining the two terms, I obtain that

\[
\sup_{s \in S_n, s \geq [np]} \left| \mathbb{E} \exp \left( \tilde{g}_1 \int_{E_1} g_1 d\tilde{N}_1' \right) - \mathbb{E} \exp \left( \tilde{g}_1 \int_{E_1} g_1 dN_1' \right) \right| \to 0.
\]

If \( s < [np] \), then

\[
\sum_{i=s}^{[np]-1} |g_1(\alpha_{1,n}(U'_{i,1} - \beta_{1,n}), X'_{i,1})| \leq \sum_{i=[np]-r_n}^{[np]-1} |g_1(\alpha_{1,n}(U'_{i,1} - \beta_{1,n}), X'_{i,1})| = o_p(1).\]

By the same argument, I have

\[
\sup_{s \in S_n, s < [np]} \left| \mathbb{E} \exp \left( \tilde{g}_1 \int_{E_1} g_1 d\tilde{N}_1' \right) - \mathbb{E} \exp \left( \tilde{g}_1 \int_{E_1} g_1 dN_1' \right) \right| \to 0.
\]

209
To sum up, I have \( \sup_{s \in S_n} \left| \mathbb{E} \exp \left( it_1 \int_{E_1} g_1 d\tilde{N}_1 \right) - \mathbb{E} \exp \left( it_0 \int_{E_0} g_1 dN_0 \right) \right| \to 0. \) By the same manner, I can show that

\[
\sup_{s \in S_n} \left| \mathbb{E} \exp \left( \tilde{t}_0 \int_{E_0} g_0 d\tilde{N}_0 \right) - \mathbb{E} \exp \left( \tilde{t}_0 \int_{E_0} g_0 dN_0 \right) \right| \to 0.
\]

This implies

\[
\sum_{s \in S_n} C_n^* p^s (1 - p)^{n-s} \mathbb{E} \exp \left( it_1 \int_{E_1} g_1 d\tilde{N}_1 \right) \mathbb{E} \exp \left( \tilde{t}_0 \int_{E_0} g_0 d\tilde{N}_0 \right) - \mathbb{E} \exp \left( \tilde{t}_0 \int_{E_0} g_0 dN_0 \right) \to 0.
\]

Combining (A.1.26) and (A.1.28),

\[
\left| \mathbb{E} \exp \left( it_1 \Psi_{1,n} + \tilde{t}_0 \Psi_{0,n} \right) - \mathbb{E} \exp \left( \tilde{t}_0 \int_{E_0} g_0 dN_0 \right) \right| \to 0.
\]

Last, notice that the random variable \( \int_{E_j} g_j dN_j \) is uniquely determined by its characteristic function

\[
\mathbb{E} \left( \exp(\tilde{t} \int_{E_j} g_j dN_j) \right) = \exp \left( - \int_{E_j} (1 - \exp(-\tilde{t} g_j)) d\mu_j \right).
\]

Similarly, the random variable \( \int_{E_j} (d + (1 - d)(1 - j)) g_j dN_j \) is uniquely determined
by its characteristic function 

\[
\mathbb{E}\exp\left(\hat{i}t \int_{E_j} (dj + (1 - d)(1 - j))g_j dN_j\right)
\]

\[
= \exp\left(-\int_{E_j} (1 - \exp(-\hat{i}tj)(dj(1 - d)g_j))\,d\mu_j\right).
\]

In addition, I have

\[
\int_{E_j} (1 - \exp(-\hat{i}t(jd + (1 - j)(1 - d))g_j))\,d\mu_j
\]

\[
= \int_{E_j} (jd + (1 - j)(1 - d))(1 - \exp(-\hat{i}tg_j))\,d\mu_j
\]

\[
= \int_{E_j} jP(x)(1 - \exp(-\hat{i}tg_j))\,d\mu_j(u, 1, x) + \int_{E_j} (1 - j)(1 - P(x))(1 - \exp(-\hat{i}tg_j))\,d\mu_j(u, 0, x)
\]

\[
= \int_{E_j} (1 - \exp(-\hat{i}tg_j))\,d\mu_j'(u, x)
\]

that is, the two characteristic functions are the same. This implies \(\int_{E_j} g_j dN_j' = \int_{E_j} (dj + (1 - d)(1 - j))g_j dN_j\). Therefore

\[
\mathbb{E}\exp\left(\hat{i} \int_{E_1} t_1g_1 dN_1'\right) \mathbb{E}\exp\left(\hat{i} \int_{E_0} t_0g_0 dN_0'\right)
\]

\[
= \mathbb{E}\exp\left(\hat{i} \int_{E_1} t_1g_1 dN_1\right) \mathbb{E}\exp\left(\hat{i} \int_{E_0} t_0(1 - d)g_0 dN_0\right)
\]

and

\[
\left|\mathbb{E}\exp\left(\hat{i}t_1\Psi_{1,n} + \hat{i}t_0\Psi_{0,n}\right) - \mathbb{E}\exp\left(\hat{i} \int_{E_1} t_1dN_1\right) \mathbb{E}\exp\left(\hat{i} \int_{E_0} t_0(1 - d)g_0 dN_0\right)\right| \to 0.
\]

For part (4), it is easy to see that \((\hat{\mathcal{Z}}_{1,n}(k), \hat{\mathcal{Z}}_{0,n}(k))\) are piece-wise constant because for instance, when \(j = 1\) and \(k - \frac{1}{P(X_h)} < \sum_{i \neq h} \frac{T_i}{P(X_i)} 1\{\alpha_n U_{i,1} < \alpha_n U_{h,1}\} < k\) for some
such that \( T_h = 1 \), then \( \hat{Z}_{1,n}(k) = \alpha_n U_{h,1} \). The discontinuity of the sample path only occurs at \( k - \frac{1}{h} P(X_h) = 1 \), \( \hat{Z}_1, n_p k_q \alpha n U_h, 1 \). The discontinuity of the sample path only occurs at \( k - \frac{1}{h} P(X_h) = 1 \), \( \hat{Z}_1, n_p k_q \alpha n U_h, 1 \). W.l.o.g., I assume 0 < \( \hat{P}(X_i) < 1 \) for all \( i \). This implies the distances between the two closest discontinuities of the sample paths are \( \min_{1 \leq i \leq n} 1 \).

### A.1.18 Proof of Lemma A.1.3

For (1), I compute its characteristic function conditioning on data \( \Phi_n \). Let \( \tilde{i} \) be the imaginary number. I have

\[
\mathbb{E} \left\{ \exp \left[ \frac{\tilde{i}t}{n} \sum_{i=1}^{n} \sum_{i=1}^{n} \mathbb{1}\{I_t = i\} W_j(D_{i, P(X_i)}) \right] | \Phi_n \right\} = \left\{ \mathbb{E} \left[ \exp \left( \frac{\tilde{i}t}{n} \sum_{i=1}^{n} \mathbb{1}\{I_1 = i\} W_j(D_{i, P(X_i)}) \right) | \Phi_n \right] \right\}^n
\]

By the Taylor expansion, \( \sum_{i=1}^{n} 1 - \exp \left( \frac{\tilde{i}t}{n} W_j(D_{i, P(X_i)}) \right) - \tilde{i}t \frac{1}{n} \sum_{i=1}^{n} W_j(D_{i, P(X_i)}) \) → 0 a.s. By SLLN,

\[
\frac{1}{n} \sum_{i=1}^{n} W_j(D_{i, P(X_i)}) \rightarrow EW_j(D_{i, P(X_i)}) = 1 \ a.s.
\]

So \( \mathbb{E} \left\{ \exp \left[ \frac{\tilde{i}t}{n} \sum_{i=1}^{n} \sum_{i=1}^{n} \mathbb{1}\{I_t = i\} W_j(D_{i, P(X_i)}) \right] | \Phi_n \right\} \rightarrow \exp(\tilde{i}t) \ a.s. \), which implies the desired result.

For (2), I first note that \( \sum_{i=1}^{n} \mathbb{1}\{\alpha_{j,n}(U_{i,j} - \beta_{j,n}) , D_i, X_i \} \sim \sum_{i=1}^{n} \mathbb{1}\{J_{i,j} , D_{i,j} , X_{i,j} \} \) by Lemma A.1.2(2). Then (2) follows by Proposition 6.3 of Resnick (2007).
For (3),

\[ E \exp \left( i(t_1 \Psi_{1,n} + t_0 \Psi_{0,n}) \right) \]

\[ = E \exp \left( \sum_{i=1}^{n} \sum_{j=1}^{n} 1 \{ I_e = i \} \tilde{i} (t_1 g_{1,n}(\alpha_{1,n} U_{1,n}, X_i, z_1) + t_0 g_{0,n}(\alpha_{0,n} U_{0,n}, X_i, z_0)) \right) \]

\[ = E \left[ \frac{1}{n} \sum_{i=1}^{n} \exp \left( \tilde{i}(t_1 D_i g_{1,n}(\alpha_{1,n} U_{1,n}, X_i, z_1) + t_0 (1 - D_i) g_{0,n}(\alpha_{0,n} U_{0,n}, X_i, z_0)) \right) \right]^n \]

\[ = E \left[ 1 - \frac{1}{n} \sum_{i=1}^{n} \left( 1 - \exp(\tilde{i}(t_1 D_i g_{1,n}(\alpha_{1,n} U_{1,n}, X_i, z_1) + t_0 (1 - D_i) g_{0,n}(\alpha_{0,n} U_{0,n}, X_i, z_0))) \right) \right]^n. \]

Conditioning on \( D_1 = \cdots = D_s = 1 \) and \( D_{s+1} = \cdots = D_n = 0 \), I have

\[ \sum_{i=1}^{n} (1 - \exp(\tilde{i}(t_1 D_i g_{1,n}(\alpha_{1,n} U_{1,n}, X_i, z_1) + t_0 (1 - D_i) g_{0,n}(\alpha_{0,n} U_{0,n}, X_i, z_0)))) \]

\[ = \sum_{i=1}^{s} (1 - \exp(\tilde{i} t_1 g_{1,n}(\alpha_{1,n} U'_{1,n}, X'_i, z_1))) + \sum_{i=s+1}^{n} (1 - \exp(\tilde{i} t_0 g_{0,n}(\alpha_{0,n} U'_{0,n}, X'_i, z_0))) \]

\[ = J_{1,s,n} + J_{0,s,n}, \]

(A.1.29)

in which \((U'_{i,j}, X'_{i,j})\) is defined in the proof of Lemma A.1.2 and \( p = P(D = 1) \). Then \( J_{1,s,n} \perp J_{0,s,n} \) and

\[ E \exp \left( i(t_1 \Psi_{1,n} + t_0 \Psi_{0,n}) \right) = \sum_{s=0}^{n} C_s p^s (1-p)^{n-s} E \left[ 1 - \frac{1}{n} (J_{1,s,n} + J_{0,s,n}) \right]^n. \]

Similar to the proof of Lemma A.1.2, it can be shown that \( J_{j,s,n} - \int (1-\exp(\tilde{i} t_j g_j)) dN'_j = \)
for any \( j \), the dominated convergence theorem because 
\[
|\frac{1}{n} (J_{1,s,n} + J_{0,s,n})^n | \leq 1.
\] The third line is because \( J_{1,s,n} \perp J_{0,s,n} \) and thus so are their limits. The fourth line is because, for any \( f \in C_K(E_j) \), \( \sum_{E_j} f dN_j = \sum_{E_j} (dj + (1-d)(1-j)) f dN_j \). The last line is because for example for \( j = 1 \) and any \( f \in C_K(E_1) \),
\[
\mathbb{E} \exp \left( \int_{E_1} f dN_1 \right) = \mathbb{E} \exp \left( \sum_{i=1}^{\infty} \Gamma_{i,1} f (\mathcal{J}_{i,1}, \mathcal{D}_{i,1}, \mathcal{X}_{i,1}) \right)
\]
that is, when \( k - \frac{1}{P(X_h^*)} < \sum_{i \neq h} \frac{D_i^*}{P(X_h^*)} \mathbb{1}\{\alpha_n U_{i,1}^* < \alpha_n U_{h,1}^*\} < k \) for some \( h \) such that \( D_h^* = 1 \), then \( \hat{Z}_{i,1}^* = \alpha_n U_{i,1}^* \). And the discontinuity of the sample path occurs at \( k - \frac{1}{P(X_h^*)} = \sum_{i \neq h} \frac{D_i^*}{P(X_h^*)} \mathbb{1}\{\alpha_n U_{i,1}^* < \alpha_n U_{h,1}^*\} \) or \( k = \sum_{i \neq h} \frac{D_i^*}{P(X_h^*)} \mathbb{1}\{\alpha_n U_{i,1}^* < \alpha_n U_{h,1}^*\} = k \). W.l.o.g., I assume \( \hat{P}(X_i) < 1 \) for all \( i \). This implies the distances between the two closest discontinuities of the sample paths are \( \min_{1 \leq i \leq n} \frac{1}{P(X_i)} \geq 1 \).

**A.1.19 Proof of Lemma A.1.4**

For (1), Let \( \tilde{t} \) be the imaginary number. When replacement is allowed,

\[
\mathbb{E}\left( \exp\left( \tilde{t} \left( \frac{1}{b} \sum_{i=1}^{n} \sum_{t=1}^{m} \mathbb{1}\{I_t = i\} W_j(D_i, P(X_i)) \right) \right) \left| \Phi_n \right. \right) \\
= \left[ \mathbb{E}\left( \exp\left( \tilde{t} \left( \frac{1}{b} \sum_{i=1}^{n} \mathbb{1}\{I_1 = i\} W_j(D_i, P(X_i)) \right) \right) \left| \Phi_n \right. \right) \right]^b \\
= \left[ \mathbb{E}\exp\left( \tilde{t}\left( \frac{1}{b} \sum_{i=1}^{n} \mathbb{1}\{I_1 = i\} W_j(D_i, P(X_i)) \right) \right) \left| \Phi_n \right. \right]^b \\
= \left[ 1 - \frac{1}{b} \sum_{i=1}^{n} \left( 1 - \exp\left( \tilde{t}\left( \frac{1}{b} W_j(D_i, P(X_i)) \right) \right) \right) \right]^b .
\]

Because \( \frac{b}{n} \left\{ \sum_{t=1}^{m} \left[ 1 - \exp\left( \tilde{t}\left( \frac{1}{b} W_j(D_i, P(X_i)) \right) \right) \right] \right\} \to \tilde{t} \) as \( b, n \to \infty \) a.s., the characteristic function converges to \( \exp(\tilde{t}) \). This implies that

\[
\frac{1}{b} \sum_{i=1}^{n} \sum_{t=1}^{m} \mathbb{1}\{I_t = i\} W_j(D_i, P(X_i)) \to 1 \text{ a.s.}
\]

When replacement is not allowed,

\[
\frac{1}{b} \sum_{i=1}^{n} P_{n,i} W_j(D_i, P(X_i)) = \frac{1}{b} \sum_{i=1}^{n} (P_{n,i} - \frac{b}{n}) W_j(D_i, P(X_i)) + \frac{1}{n} \sum_{i=1}^{n} W_j(D_i, P(X_i)).
\]

(A.1.31)
The second term of (A.1.31) converges to 1 almost surely by SLLN. For the first term of
(A.1.31), \( W_j \) is bounded and \( \mathbb{E}(\frac{1}{b} \sum_{i=1}^{n} (P_{n,i} - \frac{b}{n}) W_j(D_i, P(X_i)))^2 \lesssim \frac{1}{b} + \frac{1}{n} \to 0. \)
This concludes part (1).

For part (2), \( E P_{n,i} = \frac{b}{n} \) and \( \hat{N}_j := \sum_{i=1}^{n} 1\{\alpha_{j,n}(U_{i,j} - \beta_{j,n}), X_i, D_i\} \to N_j \) by
Proposition 6.2 of Resnick (2007), for \( \hat{N}_j^* \) and \( N_j \) as random element in the space of
point measure,

\[
P(\hat{N}_j^* \in \{(\alpha_{j,n}(U_{i,j} - \beta_{j,n}), X_i, D_i\}_{i=1}^{n}) \overset{p}{\rightarrow} P(N_j \in \cdot).
\]

Taking expectation on both sides, I obtain \( \hat{N}_j^* \to N_j \).

For part (3), I first denote \((U_{i,j}', X_{i,j}')\) as is defined in the proof of Lemma A.1.2
and \( p = P(D = 1). \) When replacement is allowed,

\[
E \exp(i(t_1 \Psi_{1,n} + t_0 \Psi_{0,n}))
= E \exp \left( i \left( \sum_{i=1}^{b} \sum_{i=1}^{n} 1\{I_i = i\} (t_1 D_i g_1(\alpha_{1,b}(U_{i,1} - \beta_{1,b}), X_i)
+ t_0 (1 - D_i) g_0(\alpha_{0,b}(U_{i,0} - \beta_{0,b}), X_i)) \right) \right)
= E \left[ 1 - \frac{1}{b} \left( \sum_{i=1}^{n} (1 - \exp \left( i \left( t_1 D_i g_1(\alpha_{1,b}(U_{i,1} - \beta_{1,b}), X_i)
+ t_0 (1 - D_i) g_0(\alpha_{0,b}(U_{i,0} - \beta_{0,b}), X_i) \right) \right) \right) \right]^b
= \sum_{s=0}^{n} C_s P^s (1 - p)^{n-s} E \left\{ 1 - \frac{1}{b} \left[ \sum_{i=1}^{s} \left( 1 - \exp \left( i t_1 D_i g_1(\alpha_{1,b}(U_{i,1} - \beta_{1,b}), X_{i}) \right) \right) \right] \right\}^b
+ \frac{b}{n} \sum_{i=s+1}^{n} \left( 1 - \exp \left( i t_0 g_0(\alpha_{0,b}(U_{i,0} - \beta_{0,b}), X_{i}) \right) \right) \right\}^b.
\]

For \( s = [np] \), \( E \frac{b}{n} \sum_{i=1}^{[np]} 1\{\alpha_{1,b}(U_{i,1}' - \beta_{1,b}), X_{i}'\} \in \cdot \to \mu_{1}'(\cdot) \) and \( E \frac{b}{n} \sum_{i=[np]+1}^{n} 1\{\alpha_{0,b}(U_{i,0}' - \beta_{0,b}), X_{i}'\} \in \cdot \to \mu_{0}'(\cdot) \), where \( \mu_{j}' \) is defined as the mean measure of \( N_j' \) and \( N_j' \) is de-
fined in Lemma A.1.2. Then by Theorem 5.3 of Resnick (2007), $\frac{b}{n} \sum_{i=1}^{[np]} I\{ (\alpha_{1,b}(U_{i,1} - \beta_{1,b}), X_i^t) \} \rightsquigarrow N'_t$ as $\frac{b}{n} \to 0$. By the same argument in the proof of (3) of Lemma A.1.2, I can show that this convergence is uniform over $|s - np| \leq r_n$. Therefore, uniformly over $|s - np| \leq r_n$,

$$
\frac{b}{n} \sum_{i=1}^{s} \left( 1 - \exp(\tilde{i}(t_1g_1(\alpha_{1,b}(U_{i,1} - \beta_{1,b}), X_i^t))) \right) \mathop{\to}^p \int_{E'_1} \left[ 1 - \exp(\tilde{i}(t_1g_1(u,x))) \right] d\mu'_1,
$$

and

$$
\frac{b}{n} \sum_{i=s+1}^{n} \left( 1 - \exp(t_0g_0(\alpha_{0,b}(U_{i,0} - \beta_{0,b}), X_i^t)) \right) \mathop{\to}^p \int_{E'_0} \left[ 1 - \exp(\tilde{i}(t_0g_0(u,x))) \right] d\mu'_0.
$$

Since the term inside the expectation of the RHS of (A.1.32) is bounded by 1, by the dominated convergence theorem, the RHS of (A.1.32) converges to

$$
\exp \left\{ \int_{E'_1} \left[ 1 - \exp(\tilde{i}(t_1g_1(u,x))) \right] d\mu'_1 + \int_{E'_0} \left[ 1 - \exp(\tilde{i}(t_0g_0(u,x))) \right] d\mu'_0 \right\}
$$

$$
= \exp \left\{ \int_{E_1} \left[ 1 - \exp(\tilde{i}(t_1g_1(u,x))) \right] d\mu_1 + \int_{E_0} \left[ 1 - \exp(\tilde{i}(t_0(1-d)g_0(u,x))) \right] d\mu_0 \right\}
$$

$$
= \mathbb{E} \exp \left( \tilde{i}t_1 \int_{E_1} dg_1dN_1 \right) \mathbb{E} \exp \left( \tilde{i}t_0 \int_{E_0} dg_0dN_0 \right),
$$

in which the first equality is by the relation between $\mu_j$ and $\mu'_j$ and the second equality is by the definition of Laplace functional of Poisson random measure with mean measure $\mu_j$.

If replacement is not allowed, then by the exchangeability of the weights $P_{n,i}$,

$$
\mathbb{E} \exp(\tilde{i}(t_1\Psi_{1,n} + t_0\Psi_{0,n}))
$$

$$
= \mathbb{E} \exp \left( \tilde{i}(t_1 \sum_{i=1}^{b} D_ig_1(\alpha_{1,b}(U_{i,1} - \beta_{1,b}), X_i) + t_0 \sum_{i=1}^{b} (1 - D_i)g_0(\alpha_{0,b}(U_{i,0} - \beta_{0,b}), X_i)) \right)
$$

$$
= \mathbb{E} \exp \left( \tilde{i}t_1 \int_{E_1} dg_1dN_1 \right) \mathbb{E} \exp \left( \tilde{i}t_0 \int_{E_0} dg_0dN_0 \right),
$$

217
in which the second equality is by the same argument in the proof of (3) in Lemma A.1.2 in which \( n \) is replaced by \( b \).

(4) holds for the same reason as in the proof of (4) in Lemma A.1.3.

A.1.20 Tightness, uniqueness and continuity

**Lemma A.1.5.** \( Z_{j,\infty}(k), j = 0, 1 \) are tight.

**Proof.** Here I focus on the case for \( j = 1 \). The proof follows the proof of Lemma 9.7 in Chernozhukov (2005a). The difference here is that \( l_\delta(u, v) \) is reweighted by the inverse propensity score \( \frac{d}{P(x)} \).

First, note that the limiting objective function is \( Q_{1,\infty}(z_1) = -kz_1 + \int_z \frac{d}{P(x)}(z_1 - j)^+dN_1(j, d, x) \) when \( j > -\delta \). I can choose \( z^f \) such that \( -kz^f + \int_z \frac{d}{P(x)}(z^f - j)^+dN_1(j, d, x) = O_p(1) \). Let \( z^* = z^f + Mv \), where \( v = \pm 1 \). Then by the convexity of objective function in \( z \) and the argument between Equation (9.74) and (9.75) of Chernozhukov (2005a), I only need to show that, for any \( K \) and \( \varepsilon > 0 \), there is an \( M \) large enough such that

\[
P(\min_{v = \pm 1} Q_{1,\infty}(z^*) > K) \geq 1 - \varepsilon. \tag{A.1.33}
\]

The claim holds trivially when \( v = -1 \). For \( v = 1 \), first note that \( P(x) \leq 1 - c \).

When \( Y_1 \) has the type 1 or 3 tail,

\[
\int_z \frac{d}{P(x)}(z^f + M - j)^+dN_1(j, d, x)
\]

\[
\geq \int_{[0, k] \times \{1\} \times \text{Supp}(x)} \frac{d}{P(x)}(z^f + M - j)^+dN_1(j, d, x)
\]

\[
\geq N([0, k] \times \{1\} \times \text{Supp}(x))(z^f + M - k)^+ \frac{1}{1 - c}.
\]

218
Because \( N([0, \kappa] \times \{1\} \times \text{Supp}(\mathcal{X})) \) is a Poisson random variable with mean

\[
\int P(x) P_1^+(dx | Y = q_1(0)) h(\kappa) \to \infty
\]
as \( \kappa \to \infty \). For \( \kappa \to \infty \), \( N([0, \kappa] \times \{1\} \times \text{Supp}(\mathcal{X})) > (k + 1)(1 - c) \) with probability greater of equal to \( 1 - \epsilon \).

When \( Y_1 \) has type 2 tail, I have, for any \( \kappa < 0 \),

\[
\int \mathbb{E} \frac{d}{P(x)} (z^f + M - j)^+ dN_1(j, d, x)
\]bones equal to 1

\[
\geq \int_{[\kappa, 1]} \frac{d}{P(x)} (z^f + M - j)^+ dN_1(j, d, x)
\]bones equal to 1

\[
\geq N([\kappa, 1] \times \{1\} \times \text{Supp}(\mathcal{X})) \frac{(z^f + M - \kappa)^+}{1 - c}.
\]

Then similarly, \( N([\kappa, 1] \times \{1\} \times \text{Supp}(\mathcal{X})) \) is a Poisson random variable with mean

\[
\int P(x) P_1^+(dx | Y = q_1(0)) h(\kappa) \to \infty \text{ as } \kappa \to 0.
\]

For \( \kappa \to 0 \), \( N([\kappa, 1] \times \{1\} \times \text{Supp}(\mathcal{X})) > (k + 1)(1 - c) \) with probability greater of equal to \( 1 - \epsilon \).

So by letting \( M \) be large enough, with probability greater or equal to \( 1 - \epsilon \), I have

\[
Q_{1, \infty}(z^*) = -kz^f - kM + \int \mathbb{E} \frac{d}{P(x)} (z^f + M - j)^+ dN_1(j, d, x)
\]bones equal to 1

\[
\geq -kz^f - kM + (z^f + M - \kappa)^+(k + 1) > K.
\]

This verifies (A.1.33).

\[\square\]

**Lemma A.1.6.** Let \( M(l) \) be the set of \( l \)-element subsets of \( N = \{1, 2, \ldots, 1\} \). For

\( j = 0, 1 \), the sequence \((D_i, \mathcal{X}_{i,j})\) are i.i.d such that \( D_i \) is Bernoulli distributed with

success probability \( P(\mathcal{X}_{i,j}) \) and \( \mathcal{X}_{i,j} \) has law \( P_j^+(X \in \cdot | Y_j = q_j(0)) \). If \( P(\sum_{i \in \mu} \frac{1}{P(\mathcal{X}_{i,j})} = k) = 0, P(\sum_{i \in \mu} \frac{1}{P(\mathcal{X}_{i,j})} + \frac{1}{P(\mathcal{X}_{h,1})} = k) = 0 \), \( P(\sum_{i \in \mu} \frac{1}{1 - P(\mathcal{X}_{i,j})} = k) = 0 \) and \( P(\sum_{i \in \mu} \frac{1}{1 - P(\mathcal{X}_{i,j})} + \frac{1}{1 - P(\mathcal{X}_{h,0})} = k) = 0 \) for any \( h \) and \( \mu \in M(l) \), \( l \leq h - 1 \), then both \( Z_{1, \infty}(k) \) and \( Z_{0, \infty}(k) \)

are unique minimizers a.s.
Proof. Here I focus on the case for $j = 1$. Following the notation in Theorem 2.4.1, $J_i = h^{-1}_i(\sum_{l=1}^{i} E_l)$. By Proposition 6.1 of Koenker (2005) and Lemma A.1.5, $Z_{1,\infty}(k) = J_h$ for some $h$ such that $T_h = 1$. Then by taking directional derivative of the objective function,

$$k - \frac{1}{P(X_h)} \leq \sum_{i \neq h} \frac{D_i}{P(X_i)} \mathbb{1}\{J_i < J_h\} \leq k.$$  \hfill (A.1.34)

Since $J_i$ is monotonic increasing,

$$P\left(\sum_{i \neq h} \frac{D_i}{P(X_i)} \mathbb{1}\{J_i < J_h\} = k\right) \leq \sum_{i \neq h, j \notin M(1), h} P\left(\sum_{i \neq h} \frac{1}{P(X_i)} = k\right) \leq 0.$$  \hfill (A.1.35)

Similarly, $P(\sum_{i \neq h} \frac{P(X_i)}{P(X_h)} \mathbb{1}\{J_i < J_h\} + \frac{1}{P(X_h)} = k) = 0$. Therefore, the inequality (A.1.34) holds strictly. This implies $Z_{1,\infty}(k)$ is the unique minimizer. \hfill \Box

Lemma A.1.7. $Z_{j,\infty}(k)$ is continuous for any $k$ and $j = 0, 1$. If $k'(m - 1) > \frac{1}{\inf_{x \in \text{Supp}(X)} P(x)}$ and $k'(m - 1) > \frac{1}{\inf_{x \notin \text{Supp}(X)} (1 - P(x))}$, then

$$\sqrt{k'} \frac{Z_{j,\infty}(k) + c}{\max(Z_{1,\infty}(mk') - Z_{1,\infty}(k'), \rho(Z_{0,\infty}(mk') - Z_{0,\infty}(k')))}$$

is also continuous for $j = 0, 1$.

Proof. $Z_{1,\infty}(k) = J_h$ for some $h$ with $T_h = 1$. Because $J_h$ is continuous, $P(Z_{1,\infty}(k) = z) = \sum_h P(J_h = z) = 0$. Therefore, $Z_{1,\infty}(k)$ is continuous. Similarly, $Z_{0,\infty}(k)$ is also continuous. Assume $h_1$ and $h_2$ solve the following two first order conditions:

$$k' - \frac{1}{P(X_{h_1})} \leq \sum_{i \neq h_1} \frac{D_i}{P(X_i)} \mathbb{1}\{J_i < J_{h_1}\} \leq k',$$
\[ mk' - \frac{1}{P(\mathcal{X}_{h_2})} \leq \sum_{i \neq h_2} \frac{D_i}{P(\mathcal{X}_i)} \mathbb{1}\{\mathcal{I}_i < \mathcal{I}_{h_2}\} \leq mk'. \]

Then \( h_1 = h_2 = h \) implies \( (m-1)k' \leq \frac{1}{P(\mathcal{X}_h)} \) for some \( \mathcal{X}_h \in \text{Supp}(\mathcal{X}) \). However, the imposed condition rules out this situation. Thus \( h_1 \neq h_2 \) and \( Z^*_j,\infty(mk') \neq Z^*_j,\infty(k') \).

In fact, following the same argument in step 3 of proof of Lemma E.1 in Chernozhukov and Fernández-Val (2011), I can prove that \( \mathbb{E}[\mathcal{I}_1,\mathcal{I}_2 | \mathcal{J}_1,\mathcal{I}_3,\mathcal{I}_4] \geq 0 \) and \( \mathbb{E}[\mathcal{I}_1,\mathcal{I}_2 | \mathcal{J}_1,\mathcal{I}_3,\mathcal{I}_4] \leq \mathbb{E}[\mathcal{I}_1,\mathcal{I}_2 | \mathcal{J}_1,\mathcal{I}_3,\mathcal{I}_4] \).

Next, I aim to show \( \sup_{k \in [\kappa_1, \kappa_2]} |Z^*_j(k)|/\sigma(k) \) is continuous. Recall the definition of \( J_{1,i} \) and \( J_{0,i} \) in Theorem 2.4.1. I rely on the next technical assumption to derive the result.

**Assumption 40.** If \( \tilde{\rho} \in (0, \infty) \), for any pair of positive integers \( (h_0, h_1) \),

\[
\left| J_{h_1,1} - \tilde{\rho}J_{h_0,0} - (\eta_1(k) - \tilde{\rho}\eta_0(k)) \right| \leq \sigma(k)
\]

has at most \( L \) local extremum which are denoted as \( \{k^*_l(J_{h_1,1} - \tilde{\rho}J_{h_0,0})\}_{l=1}^L \) for some finite integer \( L \). Furthermore, the following two conditions holds:

1. \( k^*_l(J_{h_1,1} - \tilde{\rho}J_{h_0,0}) \) is continuously distributed for \( l = 1, \cdots, L \).

2. For any positive integers \( (h_0, h_1) \), any \( z \), and any \( l = 1, \cdots, L \),

\[
P \left( \left| J_{h_1,1} - \tilde{\rho}J_{h_0,0} - (\eta_1(k) - \tilde{\rho}\eta_0(k)) \right| = z \middle| k^*_l(J_{h_1,1} - \tilde{\rho}J_{h_0,0}) = k \right) = 0
\]

for almost all \( k \in [\kappa_1, \kappa_2] \).

If \( \tilde{\rho} = 0 \), for any pair of positive integers \( (h_0, h_1) \), \( |J_{h_1,1} - \eta_1(k)|/\sigma(k) \) has at most \( L \) local extremum which are denoted as \( \{k^*_l(J_{h_1,1})\}_{l=1}^L \) for some finite integer \( L \). Furthermore, the following two conditions holds:
1. \( k_1^* (J_{h_1,1}) \) is continuously distributed for \( l = 1, \cdots, L \).

2. For any positive integers \((h_0, h_1)\), any \( z \), and any \( l = 1, \cdots, L \),

\[
P \left( \left| \frac{J_{h_1,1} - \eta_1(k)}{\sigma(k)} \right| = z \mid k_1^* (J_{h_1,1} - \tilde{\rho} J_{h_0,0}) = k \right) = 0
\]

for almost all \( k \in [\kappa_1, \kappa_2] \).

If \( \tilde{\rho} = \infty \), for any pair of positive integers \((h_0, h_1)\), \( \left| \frac{J_{h_0,0} - \eta_0(k)}{\sigma(k)} \right| \) has at most \( L \) local extremum which are denoted as \( \{k_1^* (J_{h_0,0})\}_{l=1}^L \) for some finite integer \( L \). Furthermore, the following two conditions holds:

1. \( k_1^* (J_{h_0,0}) \) is continuously distributed for \( l = 1, \cdots, L \).

2. For any positive integers \((h_0, h_1)\), any \( z \), and any \( l = 1, \cdots, L \),

\[
P \left( \left| \frac{J_{h_0,0} - \eta_0(k)}{\sigma(k)} \right| = z \mid k_1^* (J_{h_1,1} - \tilde{\rho} J_{h_0,0}) = k \right) = 0
\]

for almost all \( k \in [\kappa_1, \kappa_2] \).

This assumption is mild. For example, if \( \sigma(k) := 1 \), then because \( J_{h_1,1} - \tilde{\rho} J_{h_0,0} \) is continuously distributed and \( k_1^* (J_{h_1,1} - \tilde{\rho} J_{h_0,0}) \) does not depends on \( J_{h_1,1} - \tilde{\rho} J_{h_0,0} \), that is, it is deterministic. Then the assumption holds automatically.

**Lemma A.1.8.** \( \kappa_1 \) and \( \kappa_2 \) are not in the discontinuity of either \( Z_{1,\infty}(k) \) and \( Z_{0,\infty}(k) \), and Assumption 40 holds. If \( \tilde{\rho} \in (0, \infty) \), then

\[
\sup_{k \in [\kappa_1, \kappa_2]} \left| (Z_{1,\infty}^c(k) - \tilde{\rho} Z_{0,\infty}^c(k)) / \sigma(k) \right|
\]

is continuous.

If \( \tilde{\rho} = 0 \), then

\[
\sup_{k \in [\kappa_1, \kappa_2]} \left| Z_{1,\infty}^c(k) / \sigma(k) \right|
\]

222
is continuous.

If $\tilde{\rho} \in (0, \infty)$, then

$$\sup_{k \in [\kappa_1, \kappa_2]} |Z_{1,\infty}^c(k) / \sigma(k)|$$

is continuous.

If $k'(m - 1) > \frac{1}{\inf_{x \in \text{Supp}(X)} P(x)}$ and $k'(m - 1) > \frac{1}{\inf_{x \in \text{Supp}(X)} (1 - P(x))}$, then

$$\sup_{k \in [\kappa_1, \kappa_2]} |Z_{1,\infty}^c(k) / \sigma(k)| = \sup_{k \in [\kappa_1, \kappa_2]} \left| \frac{\sqrt{\rho}}{\sigma(k)} \frac{Z_{1,\infty}^c(k) - \tilde{\rho}Z_{0,\infty}^c(k)}{\max(Z_{1,\infty}(mk') - Z_{1,\infty}(k'), \tilde{\rho}(Z_{0,\infty}(mk') - Z_{0,\infty}^c(k'))} \right|$$

is also continuous.

**Proof.** I only consider the case for $\tilde{\rho} \in (0, \infty)$. The other two cases can be proved similarly. Let $L_{h,1} = \{k : D_h = 1, k = \sum_{i \leq h} W_1(D_{i,1}, X_{i,1}) \text{ or } k = \sum_{i \leq h} W_1(D_{i,1}, X_{i,1}) \}$ and $L_{h,0} = \{k : D_h = 1, k = \sum_{i \leq h} W_0(D_{i,0}, X_{i,0}) \text{ or } k = \sum_{i \leq h} W_0(D_{i,0}, X_{i,0}) \}$. Then the discontinuities of the sample path of $Z_{j,\infty}(k)$ is $\cup_{h \geq 1} L_{j,h}$. Since the closest distance between two distinct discontinuities of $Z_{j,\infty}(k)$ is at least 1, there are at most finite number of discontinuities of either $Z_{1,\infty}(k)$ or $Z_{0,\infty}(k)$. This implies the closest distance between two distinct discontinuities of $Z_{\infty}^c(k)$ is strictly positive. For a fixed event $\omega$, if $\sup_{k \in [\kappa_1, \kappa_2]} |Z_{\infty}^c(k)(\omega)| = z$, then there exists a convergent sequence $\hat{k}_m(\omega)^4$ with limit $\hat{k}(\omega)$ such that $|Z_{\infty}^c(\hat{k}_m(\omega))(\omega)| \to z$. Since $Z_{j,\infty}(k)$ is piece-wise constant, $\kappa_1$ and $\kappa_2$ are not in $\cup_{j=0,1} \cup_{h \geq 1} L_{j,h}$, there exist $M(\omega)$ large enough such that for $m > M(\omega)$,

$$z = \sup_{k \in [\kappa_1, \kappa_2]} |Z_{\infty}^c(k)(\omega)/\sigma(k)|$$

$$= |(Z_{1,\infty}(\hat{k}_m) - \tilde{\rho}Z_{0,\infty}(\hat{k}_m) - (\eta_1(\hat{k}) - \tilde{\rho}\eta_0(\hat{k}))/\sigma(\hat{k})|$$

$$= |(J_{h,1} - \tilde{\rho}J_{h,0} - (\eta_1(\hat{k}) - \tilde{\rho}\eta_0(\hat{k}))/\sigma(\hat{k})|,$$

$^4 \hat{k}_m(\omega)$ depends on the sample path and thus is random.
in which \( \hat{k}_m - \frac{1}{P(\hat{X}_{h,0})} < \sum_{i<h_1} W_1(D_{i,1}, X_{i,1}) < \hat{k}_m, \hat{k}_m - \frac{1}{1 - P(\hat{X}_{h,0})} < \sum_{i<h_0} W_0(D_{i,0}, X_{i,0}) < \hat{k}_m \), and \( \hat{k} \in \mathcal{L}(\hat{h}_1, \hat{h}_0) := \mathcal{L}_{h_1,1} \cup \mathcal{L}_{h_0,0} \cup \{ k_i^* (\mathcal{J}_{h_1,1} - \tilde{\rho} \mathcal{J}_{h_0,0}) \}_{i=1}^L \cup \{ \kappa_1 \} \cup \{ \kappa_2 \} \). Furthermore, let \( \mathcal{A}_h = \{ \sum_{i<h} D_{i,1} > \kappa_2, \sum_{i<h_1} (1 - D_{i,0}) > \kappa_2 \} \). Then on \( \mathcal{A}_h, \hat{h}_j \leq h \) for \( j = 0, 1 \). Therefore,

\[
P \left( \sup_{k \in [\kappa_1, \kappa_2]} |Z^n(k)(\omega)/\sigma(k)| = z \right) 
\leq \sum_{h > \kappa_2} P \left( |(\mathcal{J}_{h_1,1} - \tilde{\rho} \mathcal{J}_{h_0,0} - (\eta_1(\hat{k}) - \tilde{\rho} \eta_0(\hat{k})))/\sigma(\hat{k})| = z, \hat{k} \in \mathcal{L}(\hat{h}_1, \hat{h}_0), \mathcal{A}_h \right) 
\leq \sum_{h > \kappa_2} \sum_{h_1 \leq h, h_0 \leq h} P \left( |(\mathcal{J}_{h_1,1} - \tilde{\rho} \mathcal{J}_{h_0,0} - (\eta_1(\hat{k}) - \tilde{\rho} \eta_0(\hat{k})))/\sigma(\hat{k})| = z, \hat{k} \in \mathcal{L}(\hat{h}_1, \hat{h}_0) \right)
\]

In order to bound the last equation, I note that \( \mathcal{J}_{h_1,1} - \tilde{\rho} \mathcal{J}_{h_0,0} - (\eta_1(k) - \tilde{\rho} \eta_0(k)) \) is continuously distributed, \( \mathcal{J}_{i,j} \) is independent of \( (D_{i,j}, X_{i,j}) \) for any realization \( (h_1, h_0) \) of \( (\hat{h}_1, \hat{h}_0) \), and \( (\mathcal{J}_{h_1,1}, \mathcal{J}_{h_0,0}) \perp \perp \mathcal{L}(h_1, h_0) \). Hence, if \( \hat{k} \in \mathcal{L}_{h_1,1} \) and for instance, \( \hat{k} = \sum_{i<h_1} W_1(D_{i,1}, X_{i,1}) \), I have

\[
P \left( |(\mathcal{J}_{h_1,1} - \tilde{\rho} \mathcal{J}_{h_0,0} - (\eta_1(\hat{k}) - \tilde{\rho} \eta_0(\hat{k}))/\sigma(\hat{k})| = z, \hat{k} = \sum_{i<h_1} W_1(D_{i,1}, X_{i,1}) \right) 
\leq \int P \left( |(\mathcal{J}_{h_1,1} - \tilde{\rho} \mathcal{J}_{h_0,0} - (\eta_1(k) - \tilde{\rho} \eta_0(k))/\sigma(k)| = z \right| \sum_{i<h_1} W_1(D_{i,1}, X_{i,1}) = k \right) 
\times dP \left( \sum_{i<h_1} W_1(D_{i,1}, X_{i,1}) \leq k \right) 
= \int P \left( |(\mathcal{J}_{h_1,1} - \tilde{\rho} \mathcal{J}_{h_0,0} - (\eta_1(k) - \tilde{\rho} \eta_0(k))| = z \right) dP \left( \sum_{i<h_1} W_1(D_{i,1}, X_{i,1}) \leq k \right) 
= 0.
\]
Similarly, if $\hat{k} \in \mathcal{L}_{h_1,1}$ and $\hat{k} = \sum_{i \in h_1} W_i(\mathbf{D}_{i,1}, \mathbf{X}_{i,1})$, 

$$P\left(\left|\mathbf{J}_{h_1,1} - \hat{\mathbf{J}}_{h_0,0} - (\eta_1(\hat{k}) - \hat{\rho} \eta_0(\hat{k}))/\sigma(\hat{k})\right| = z, \hat{k} = \sum_{i \in h_1} W_i(\mathbf{D}_{i,1}, \mathbf{X}_{i,1})\right) = 0.$$ 

If $\hat{k} \in \{k^*_l(\mathbf{J}_{h_1,1} - \hat{\mathbf{J}}_{h_0,0})\}_{l=1}^L$, 

$$P\left(\left|\mathbf{J}_{h_1,1} - \hat{\mathbf{J}}_{h_0,0} - (\eta_1(\hat{k}) - \hat{\rho} \eta_0(\hat{k}))/\sigma(\hat{k})\right| = z, \hat{k} \in \{k^*_l(\mathbf{J}_{h_1,1} - \hat{\mathbf{J}}_{h_0,0})\}_{l=1}^L\right)$$

$$\leq \sum_{l=1}^L \int_{k_1}^{k_2} P\left(\left|\mathbf{J}_{h_1,1} - \hat{\mathbf{J}}_{h_0,0} - (\eta_1(k) - \hat{\rho} \eta_0(k))/\sigma(k)\right| = z \left| k^*_l(\mathbf{J}_{h_1,1} - \hat{\mathbf{J}}_{h_0,0}) = k\right.\right)$$

$$\times dP\left(k^*_l(\mathbf{J}_{h_1,1} - \hat{\mathbf{J}}_{h_0,0}) \leq k\right) = 0.$$

Last, if $\hat{k} = \kappa_1$ or $\kappa_2$, 

$$P\left(\left|\mathbf{J}_{h_1,1} - \hat{\mathbf{J}}_{h_0,0} - (\eta_1(\hat{k}) - \hat{\rho} \eta_0(\hat{k}))/\sigma(\hat{k})\right| = z, \hat{k} = \kappa_1 \text{ or } \kappa_2\right) = 0.$$ 

To sum up, I have 

$$P\left(\left|\mathbf{J}_{h_1,1} - \hat{\mathbf{J}}_{h_0,0} - (\eta_1(\hat{k}) - \hat{\rho} \eta_0(\hat{k}))/\sigma(\hat{k})\right| = z, \hat{k} \in \mathcal{L}(\hat{h}_1, \hat{h}_0)\right) = 0.$$ 

Hence the first result follows. The second result can be proved in a same manner as in Lemma A.1.7. \qed
Appendix B

Appendix for Chapter 3

B.1 Details on the data-driven \( \tau_n \)

We provide in this section a rationale for the construction of the data-driven \( \tau_n \) detailed in Section 3.3. We study for that purpose the asymptotic behavior of \( \hat{\delta} \) for sequences \( \tau' \) that do not satisfy Assumption 27 (iii), but only \( \sqrt{\tau_n n f(\gamma(\tau_n))} = O(1) \).

We show that in this case, \( \hat{\delta} \) has an asymptotic bias. Then we relate this bias with the asymptotic behavior of the \( J \) test statistic \( T_J(\tau'_n) \), and show how this can be used to select a quantile index for which the asymptotic bias is small.

First, let us define

\[
\mu(\tau) = \frac{E[(\tau - 1\{\min \{Y \leq \gamma(\tau) + X'\beta(\tau)\} \mid X] }{\tau} = \frac{E[(\tau - 1\{\min \{D \leq \min \{Q_{\tau}(\tau / h) + X'\delta\} \mid X] }{\tau}.
\]

As shown in the proof of Theorem 3.3.1, \( \mu(\tau) \) is the core component of the bias induced by the fact that (3.3.2) is an equivalence instead of an equality. \( f(\gamma(\tau_n)) \) in Assumption 27 can then be viewed as an envelope of \( \mu(\tau) \). Under Assumption 27-(iii), \( \sqrt{\tau_n n} \mu(\tau_n) \to 0 \), meaning that the asymptotic bias vanishes. In what follows, we derive the asymptotic bias of our estimator \( \hat{\delta} \) as a function of \( \mu(\tau) \) and propose
a subsampling method to approximate this bias.

From (B.3.11) and the linear representation of $\hat{Z}_n(1)$ below (B.3.14) in the proof of Theorem 3.3.1, we have, for any sequence $\tau_n$ that satisfies Assumption 27,

$$\sqrt{\tau_n}n(\hat{\delta} - \delta) = \log(m)(G'W^*_\delta G)^{-1}G'W^*_\delta \alpha_n(\tau_n)g_n(\delta) + o_P(1),$$

where $\alpha_n(\tau) = \sqrt{\tau_n}/(\gamma(m\tau) - \gamma(\tau))$. We also show in the proof of Theorem 3.3.1 that

$$\alpha_n(\tau_n)g_n(\delta) = (I \otimes \Delta)\Gamma \hat{Z}_n(\tau_n),$$

where $\hat{Z}_n(\tau_n)$ is asymptotically normal with mean 0 when the asymptotic bias of $\hat{\delta}$ is zero. In order to analyze situations where a sequence $\tau'_n$ only satisfies $\sqrt{\tau_n}nf(\gamma(\tau'_n)) = O(1)$, consider

$$\hat{Z}_n(\tau) = \log(m)\hat{Z}_n(\tau) + Q^{-1}_H\sqrt{\tau_n}b(\tau),$$

with $b(\tau) = (\mu(\tau), \sqrt{l_1\mu(l_1\tau)}, \ldots, \sqrt{l_J\mu(l_J\tau)})'$. Then one can show that

$$\sqrt{\tau_n}n(\hat{\delta} - \delta) = (G'W^*_\delta G)^{-1}G'W^*_\delta (I_J \otimes \Delta)\Gamma \hat{Z}_n(\tau'_n)$$

$$- (G'W^*_\delta G)^{-1}G'W^*_\delta (I_J \otimes \Delta)(I_{J+1} \otimes Q^{-1}_H)\sqrt{\tau_n}nB(\tau'_n) + o_P(1).$$

$\hat{Z}_n(\tau'_n)$ is asymptotically normal with mean 0 by definition of $b(\tau)$. Hence, the second term is the asymptotic bias of $\hat{\delta}$. We seek to approximate the norm of this bias. In order to do so, we consider the minimum distance statistic, which is commonly used to conduct a specification test in the context of minimum distance estimation. By plugging in the minimum distance estimator $\hat{\delta}$, we obtain

$$\log(m)\alpha_n(\tau'_n)W_{\delta}^{1/2}g_n(\hat{\delta})$$

$$= (I_{Jd} - W_{\delta}^{1/2}G(G'W^*_\delta G)^{-1}G'W_{\delta}^{1/2}) (I \otimes \Delta)\Gamma \hat{Z}_n(\tau'_n) - \sqrt{\tau_n}nB(\tau'_n) + o_P(1),$$

where $B(\tau) = (I \otimes \Delta)(I_{J+1} \otimes Q^{-1}_H)b(\tau)$. $B(\tau)$ is the bias associated with the choice of quantile index $\tau$. It follows that the J-statistic defined in Section 3.3 can be written as
This equation shows that the J-statistic on the left-hand side converges to a chi square distribution with \((J - 1)d\) degrees of freedom, plus a bias term. If \(\sqrt{\tau_n n B(\tau_n')} \to 0\), then the median of the J-statistic is asymptotically the median \(M_{(J-1)d}\) of a \(\chi^2((J - 1)d)\). On the other hand, if the asymptotic bias \(\sqrt{\tau_n n B(\tau_n')}\) does not vanish, the difference between the median of the J-statistic and \(M_{(J-1)d}\) will generally be asymptotically different from zero. Following this idea, we estimate the difference between the two medians and use it as a proxy for the asymptotic bias of \(\delta\). As indicated in the text, we rely for that purpose on subsampling.

B.2 Data appendix

We construct our NLSY97 dataset based on the interviews that were conducted during the years 2007 and 2008, using data on males from the cross-sectional sample and the oversample of blacks and Hispanics of the NLSY97. Our sample consists of the respondents who reported wages for at least one of these two years, along with the respondents who reported not working in either year (nonparticipants). Respondents with a missing AFQT score are excluded from the analysis. For the individuals working in both years, the wage variable is defined as the average of the hourly wages corresponding to the main job at the time of the interview. For those working during one year only, we define the wage variable as the hourly wage corresponding to the main job at the time of the interview in that year. Finally, we trim the data by dropping the wage observations below 1 dollar and above 118.95 dollars (corresponding to 75 dollars in 1991). We report in Table B.1 below some
descriptives corresponding to our NLSY79 and NLSY97 samples restricted to the respondents who took the ASVAB test when they were 16 or 17. Table B.2 reports the labor force participation rates for the NLSY79 and NLSY97 samples, separately for blacks and whites.

Table B.1: Descriptive statistics for the subsample with restricted age

<table>
<thead>
<tr>
<th></th>
<th>NLSY79</th>
<th></th>
<th>NLSY97</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Blacks</td>
<td>Hispanics</td>
<td>Whites</td>
<td>Blacks</td>
</tr>
<tr>
<td>AFQT</td>
<td>-0.716</td>
<td>-0.314</td>
<td>0.387</td>
<td>-0.726</td>
</tr>
<tr>
<td>Std.dev.</td>
<td>(0.812)</td>
<td>(0.935)</td>
<td>(0.966)</td>
<td>(1.037)</td>
</tr>
<tr>
<td>Std.dev.</td>
<td>(3.927)</td>
<td>(3.953)</td>
<td>(3.691)</td>
<td>(4.811)</td>
</tr>
<tr>
<td>Mother high school graduate</td>
<td>0.447</td>
<td>0.243</td>
<td>0.715</td>
<td>0.707</td>
</tr>
<tr>
<td>Father high school graduate</td>
<td>0.368</td>
<td>0.284</td>
<td>0.665</td>
<td>0.518</td>
</tr>
<tr>
<td>Mother college graduate</td>
<td>0.040</td>
<td>0.027</td>
<td>0.093</td>
<td>0.107</td>
</tr>
<tr>
<td>Father college graduate</td>
<td>0.046</td>
<td>0.050</td>
<td>0.188</td>
<td>0.086</td>
</tr>
<tr>
<td>Both parents at age 14</td>
<td>0.486</td>
<td>0.599</td>
<td>0.760</td>
<td>0.264</td>
</tr>
</tbody>
</table>

Note: Samples restricted to males. Blacks account for 31% (25%) of the NLSY79 (NLSY97) sample, while Hispanics account for 20% (21%) of the NLSY79 (NLSY97) samples.

Table B.2: Labor force participation rates (males)

<table>
<thead>
<tr>
<th></th>
<th>Blacks</th>
<th>Whites</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLSY79 full sample</td>
<td>91.02%</td>
<td>97.52%</td>
</tr>
<tr>
<td>NLSY79 with age restriction</td>
<td>90.58%</td>
<td>98.10%</td>
</tr>
<tr>
<td>NLSY97 with age restriction</td>
<td>81.43%</td>
<td>93.09%</td>
</tr>
</tbody>
</table>

B.3 Proofs of the results

In the following, we let, for any random variable $U$ and with a slight abuse of notations, $S_U^- = 1 - F_U^-$. We also let $	ilde{U} = -U$ and define $\tilde{\varepsilon} = \tilde{Y} + X'\beta$. Finally, we take the convention that intervals $[a,b]$ refer to $[b,a]$ when $b < a$, and similarly for open or semi-open intervals.
B.3.1 Proof of Theorem 3.2.1

Identification of \( \psi(.) \) and \( \sigma(.) \) follows directly from Theorem 2.1 in D'Haultfoeuille and Maurel (2013). Note that in D'Haultfoeuille and Maurel (2013) we use \( E(\exp(\lambda \varepsilon)) < +\infty \) for some \( \lambda > 0 \) instead of the weaker condition that \( S_{\exp(\varepsilon)} \) is slowly varying. An inspection of the proof reveals however that the proof only relies on the latter condition.

Turning to \( \Delta_{j,\tau} \), remark first that by independence between \( X \) and \( \varepsilon \),

\[
\Delta_{j,\tau}(x) = \partial_j \psi(x) + \partial_j \sigma(x) Q_{\varepsilon}(\tau). \tag{B.3.1}
\]

It suffices therefore to obtain bounds on \( Q_{\varepsilon}(\tau) \). We suppose hereafter that \( \partial_j \sigma(x) > 0 \). The reasoning is similar for \( \partial_j \sigma(x) < 0 \), while \( \Delta_{j,\tau}(x) \) is identified from (B.3.1) if \( \partial_j \sigma(x) = 0 \). First, we have

\[
f_D|X(u|x) = P(D = 1|X = x)f_{D=1,X}(u|x) + P(D = 0|X = x)f_{D=0,X}(u|x).
\]

Thus, for all \( x \) in the support of \( X \),

\[
f_{D|X}(u|x) \geq P(D = 1|X = x)f_{D=1,X}(u|x).
\]

By independence between \( X \) and \( \varepsilon \),

\[
f_{\varepsilon}(u) = \sup_{x \in \text{Supp}(X)} f_{D|X}(u|x) \geq \sup_{x \in \text{Supp}(X)} P(D = 1|X = x)f_{D=1,X}(u|x). \tag{B.3.2}
\]

Integrating (B.3.2) between \( -\infty \) and \( v \) implies that \( F_{\varepsilon}(v) \geq F_\varepsilon(v) \). Hence, \( Q_{\varepsilon}(\tau) \leq F_{\varepsilon}^*(\tau) \). This yields the upper bound on \( \Delta_{j,\tau}(x) \). Now, integrating (B.3.2) between \( v \) and \( +\infty \) implies that

\[
1 - F_{\varepsilon}(v) \geq \int_{v}^{+\infty} \left[ \sup_{x \in \text{Supp}(X)} P(D = 1|X = x)f_{D=1,X}(u|x) \right] du.
\]

230
Hence, \( F_\varepsilon(v) \leq 1 - F_\varepsilon(+\infty) + F_\varepsilon(v) \). As a result,

\[
Q_\varepsilon(\tau) \geq [1 - F_\varepsilon(+\infty) + F_\varepsilon]^{-1}(\tau) = F_\varepsilon^{-1}(\tau - (1 - F_\varepsilon(\varepsilon)))
\]

The lower bound on \( \Delta_{j\tau}(x) \) follows.

Now, let us show that these bounds are sharp under (3.2.2). For that purpose, we exhibit conditional cdfs \( \tilde{F}_{\varepsilon|D=0,X}(\cdot|x) \), different in general from the true ones, which rationalize the bounds and satisfy the restrictions imposed by Assumptions 19, 20(i)-(iii), 21 and (3.2.2). Note that the condition 24-(iv) only depends on the observed data and therefore needs not be verified. Note also that we can restrict to the case where \( P(D = 0|X = x) > 0 \) for almost all \( x \). For in the case where \( P(D = 0|X = x) = 0 \), Inequality (B.3.2) is actually an equality, and the two bounds coincide. The bounds then correspond to the true model, and are therefore sharp.

Now, consider the upper bound. Let \( u_0 \) be such that \( F_\varepsilon(u_0) > \tau \) and suppose that

\[
\tilde{F}_{\varepsilon|D=0,X}(u|x) = \frac{F_\varepsilon(u) - P(\varepsilon \leq u, D = 1|X = x)}{P(D = 0|X = x)}1\{u < u_0\} + \tilde{F}_{\varepsilon|D=0,X}(u|x)1\{u \geq u_0\}.
\]

Let us first show that for all \( x \), \( \tilde{F}_{\varepsilon|D=0,X}(\cdot|x) \) is indeed a cdf. It suffices to show that its limit at \(-\infty\) is zero, that it is increasing and right-continuous on \((-\infty, u_0)\) and \( \lim_{u \uparrow u_0} \tilde{F}_{\varepsilon|D=0,X}(u|x) \leq F_{\varepsilon|D=0,X}(u_0|x) \). The first point holds because \( \lim_{u \to -\infty} F_\varepsilon(u) = \lim_{u \to -\infty} P(\varepsilon \leq u, D = 1|X = x) = 0 \). The second point follows by remarking that

\[
F_\varepsilon(v) - P(\varepsilon \leq v, D = 1|X = x) = \int_{-\infty}^v \left\{ \sup_{x' \in \text{Supp}(X)} [f_{\varepsilon|D=1,X}(u|x')P(D = 1|X = x')] - f_{\varepsilon|D=1,X}(u|x)P(D = 1|X = x) \right\} du.
\]

The integral form implies that \( \tilde{F}_{\varepsilon|D=0,X}(\cdot|x) \) is right-continuous. Because the term in braces is positive, \( \tilde{F}_{\varepsilon|D=0,X}(\cdot|x) \) is also increasing. Finally, the third point follows
because for any $u$,

$$\frac{F_\varepsilon(u) - P(\varepsilon \leq u, D = 1|X = x)}{P(D = 0|X = x)} \leq \frac{F_\varepsilon(u) - P(\varepsilon \leq u, D = 1|X = x)}{P(D = 0|X = x)} = F_{\varepsilon|D=0,X}(u|x).$$

Now, let us prove that the conditional cdfs $\tilde{F}_{\varepsilon|D=0,X}(\cdot|\cdot)$ rationalize the bounds and satisfy the restrictions of the model. First,

$$\tilde{F}_{\varepsilon|X}(u|x) = \tilde{F}_{\varepsilon|D=0,X}(u|x)P(D = 0|X = x) + P(\varepsilon \leq u, D = 1|X = x)$$

$$= F_\varepsilon(u)1\{u < u_0\} + F_\varepsilon(u)1\{u \geq u_0\}. \quad (B.3.3)$$

The right-hand side does not depend on $x$. Therefore, $\tilde{F}_{\varepsilon|D=0,X}$ satisfies Assumption 19. $(B.3.3)$ also implies that for any $\tau' \leq \tau$,

$$\tilde{F}_\varepsilon^\tau(\tau') = F_\varepsilon^\tau(\tau').$$

Therefore, the conditional cdfs $\tilde{F}_{\varepsilon|D=0,X}(\cdot|\cdot)$ rationalize $\Xi_{ijr}$. Now, because $\tilde{f}_\varepsilon(u)$ is equal to the true $f_\varepsilon(u)$ for $u$ large enough, the conditional cdfs $\tilde{F}_{\varepsilon|D=0,X}(\cdot|\cdot)$ satisfy Assumption 20. Similarly, by Bayes' theorem, we have, for $y$ large enough,

$$\tilde{P}(D = 1|X = x, Y^* = y) = \frac{f_{Y|D=1,X=x}(y|x)P(D = 1|X = x)}{\tilde{f}_{Y^*|X}(y|x)}$$

$$= \frac{\sigma(x)f_{Y|D=1,X=x}(y|x)P(D = 1|X = x)}{\tilde{f}_\varepsilon([(y - \psi(x))/\sigma(x)])}$$

$$= \frac{\sigma(x)f_{Y|D=1,X=x}(y|x)P(D = 1|X = x)}{f_\varepsilon([(y - \psi(x))/\sigma(x)])}$$

$$= \sigma(x)f_{Y|D=1,X=x}(y|x)P(D = 1|X = x)$$

$$= P(D = 1|X = x, Y^* = y),$$

and therefore, Assumption 21 is satisfied. This equality also ensures that (3.2.2) is satisfied. Hence, the upper bound is sharp.
Now, let us turn to the lower bound. Let $u_1$ be such that $F_x(u_1) < \tau$ and consider

$$
\tilde{F}_{\varepsilon|D=0,X}(u|x) = F_{\varepsilon|D=0,X}(u|x) 1\{u < u_1\} \\
+ \frac{1 - F_x(\varepsilon + \infty) + F_x(u) - P(\varepsilon \leq u, D = 1|X = x)}{P(D = 0|X = x)} 1\{u \geq u_1\}.
$$

(B.3.4)

As previously, $\tilde{F}_{\varepsilon|D=0,X}(\cdot|x)$ is indeed a cdf and

$$
\tilde{F}_{\varepsilon|X}(u|x) = F_x(u) 1\{u < u_1\} + [1 - F_x(\varepsilon + \infty) + F_x(u)] 1\{u \geq u_1\},
$$

so that Assumption 19 holds and $\tilde{F}_{\varepsilon|D=0,X}(\cdot)$ rationalizes the lower bound. We now check Assumption 20. For $u$ large enough,

$$
\tilde{f}_\varepsilon(u) = \sup_{x \in \text{Supp}(X)} P(D = 1|X = x) f_{\varepsilon|D=1,X}(u|x) \\
= \sup_{x \in \text{Supp}(X)} P(D = 1|X = x, \varepsilon = u) f_{\varepsilon|X}(u|x) \\
= f_\varepsilon(u) \left[ \sup_{x \in \text{Supp}(X)} P(D = 1|X = x, Y^* = \psi(x) + \sigma(x)u) \right].
$$

(B.3.5)

We now prove that $\tilde{f}_\varepsilon(u) \sim h f_\varepsilon(u)$ as $u \to \infty$. Fix $\eta > 0$. Because Supp($X$) is compact, there exists $(x_1, \ldots, x_k) \in \text{Supp}(X)^k$ such that for all $x \in \text{Supp}(X)$, $\min_{j=1\ldots,k} \|x - x_j\| < \eta$. There exists also $y_0$ such that for all $y \geq y_0$,

$$
\max_{j=1\ldots,k} |P(D = 1|X = x_j, Y^* = y) - h| < \eta.
$$

(B.3.6)

By compactness of Supp($X$) once more, there exists $u_0$ such that for all $u \geq u_0$, 

$$
\inf_{x \in \text{Supp}(X)} \psi(x) + \sigma(x)u \geq y_0.
$$

(B.3.7)
Then, for all \( x \in \text{Supp}(X) \), and all \( u \geq u_0 \),
\[
|P(D = 1|X = x, Y^* = \psi(x) + \sigma(x)u) - h| \\
\leq |P(D = 1|X = x, Y^* = \psi(x) + \sigma(x)u) - P(D = 1|X = x_j, Y^* = \psi(x) + \sigma(x)u)| \\
+ |P(D = 1|X = x_j, Y^* = \psi(x) + \sigma(x)u) - h| \\
\leq K\|x - x_j\| + \eta,
\]
where the second inequality follows by (3.2.2), (B.3.6) and (B.3.7). Choosing \( j \) such that \( \|x - x_j\| < \eta \) finally yields
\[
\sup_{x \in \text{Supp}(X)} |P(D = 1|X = x, Y^* = \psi(x) + \sigma(x)u) - h| < (K + 1)\eta.
\]
As a result,
\[
\lim_{u \to \infty} \sup_{x \in \text{Supp}(X)} P(D = 1|X = x, Y^* = \psi(x) + \sigma(x)u) = h.
\]
Hence, by (B.3.5), as \( u \to \infty \),
\[
\tilde{f}_\epsilon(u) \sim h\tilde{f}_\epsilon(u).
\]
This implies that Assumption 20-(i) holds. Now, suppose that \( \tilde{S}_{\exp(\epsilon)} \) is slowly varying. Then for all \( l > 0 \), \( \tilde{S}_{\exp(\epsilon)}(lu)/\tilde{S}_{\exp(\epsilon)}(u) \to 1 \). Now, (B.3.8) also implies that for any \( l > 0 \),
\[
\frac{S_{\exp(\epsilon)}(lu)}{S_{\exp(\epsilon)}(u)} \sim \frac{\tilde{S}_{\exp(\epsilon)}(lu)}{\tilde{S}_{\exp(\epsilon)}(u)}.
\]
This implies that \( S_{\exp(\epsilon)} \) is also slowly varying, a contradiction. Thus, Assumption 20-(ii) is satisfied. By (B.3.8) once more, there exists \( \eta > 0 \) arbitrarily small such that for all \( u \) large enough,
\[
(h - \eta)S_\epsilon(u) \leq \tilde{S}_\epsilon(u) \leq (h + \eta)S_\epsilon(u).
\]
234
Now, fix $\tau$ small enough and let $u = \tilde{S}_\epsilon^u(\tau)$. $\tilde{S}_\epsilon(u) \geq \tau$ implies $S_\epsilon(u) \geq \tau/(h + \eta)$, which yields in turn $u \leq S_\epsilon^u(\tau/(h + \eta))$. Hence, we obtain

$$\tilde{S}_\epsilon^u(\tau) \leq S_\epsilon^u(\tau/(h + \eta)) = -Q_\epsilon(\tau/(h + \eta)).$$

Now, let $u' > u$, so that $\tilde{S}_\epsilon(u') \leq \tau$. Then $u' \geq S_\epsilon^u(\tau/(h - \eta))$. Letting $u'$ tend to $u$ yields

$$\tilde{S}_\epsilon^u(\tau) \geq S_\epsilon^u(\tau/(h - \eta)) = -Q_\epsilon(\tau/(h - \eta)).$$

As a result, for any fixed $m > 1$ and letting $e = \exp(1)$,

$$\frac{\tilde{S}_\epsilon^u(m\tau) - \tilde{S}_\epsilon^u(\tau)}{S_\epsilon^u(e\tau) - \tilde{S}_\epsilon^u(\tau)} \leq \frac{Q_\epsilon(m\tau/(h - \eta)) - Q_\epsilon(\tau/(h + \eta))}{Q_\epsilon(e\tau/(h + \eta)) - Q_\epsilon(\tau/(h - \eta))}.$$

By Lemma D.2 in Appendix D, the right-hand side converges to $\log(m(h + \eta)/(h - \eta))/\log(e(h - \eta)/(h + \eta))$. Reasoning similarly on the lower bound, we obtain,

$$\log(m(h - \eta)/(h + \eta))/\log(e(h + \eta)/(h - \eta)) \leq \lim inf_{\tau \to 0} \frac{\tilde{S}_\epsilon^u(m\tau) - \tilde{S}_\epsilon^u(\tau)}{S_\epsilon^u(e\tau) - \tilde{S}_\epsilon^u(\tau)} \leq \lim sup_{\tau \to 0} \frac{\tilde{S}_\epsilon^u(m\tau) - \tilde{S}_\epsilon^u(\tau)}{S_\epsilon^u(e\tau) - \tilde{S}_\epsilon^u(\tau)} \leq \log(m(h + \eta)/(h - \eta))/\log(e(h - \eta)/(h + \eta)).$$

Because $\eta$ was arbitrary, we can make it tend to zero, thus obtaining

$$\lim_{\tau \to 0} \frac{\tilde{S}_\epsilon^u(m\tau) - \tilde{S}_\epsilon^u(\tau)}{S_\epsilon^u(e\tau) - \tilde{S}_\epsilon^u(\tau)} = \log(m).$$

This proves (see (Resnick, 1987), Proposition 0.10) that $\tilde{S}_\epsilon$ belongs to the domain of attraction of the Gumbel distribution. Hence, Assumption 20-(iii) holds.
Turning to Assumption 21, we reason as for the upper bound:

\[ \tilde{P}(D = 1|X = x, Y^* = y) = \frac{\sigma(x)f_y|D=1,X=x(y|x)P(D = 1|X = x)}{f_x[(y - \psi(x))/\sigma(x)]} \approx \frac{\sigma(x)f_y|D=1,X=x(y|x)P(D = 1|X = x)}{h f_x[(y - \psi(x))/\sigma(x)]} \approx \frac{P(D = 1|X = x, Y^* = y)}{h}. \]

Therefore, the conditional cdfs \( \tilde{F}_{d|D=0,X} \) satisfy Assumption 21, with a limit equal to 1 instead of \( h \). The result follows.

Finally, let us check (3.2.2). We have by what precedes, for \( y \) large enough

\[ \tilde{P}(D = 1|X = x, Y^* = y) = \frac{P(D = 1|X = x, Y^* = y)}{\sup_{x'\in\text{Supp}(X)} P(D = 1|X = x', Y^* = y)}. \]

Moreover, we have proved that the denominator tends to \( h \) as \( y \to \infty \). Therefore, because the true distribution satisfies (3.2.2), we have, for all \((x, x') \in \text{Supp}(X)^2 \) and all \( y \) large enough,

\[ |\tilde{P}(D = 1|X = x', Y^* = y) - \tilde{P}(D = 1|X = x, Y^* = y)| \leq \frac{K}{h - \eta} \|x - x'\|, \]

for some \( 0 < \eta < h \). This ensures that \( \tilde{F}_{d|D=0,X} \) satisfies (3.2.2), and thus that the lower bound is sharp.

**B.3.2 Derivation of the outer bounds (3.2.3)**

We only consider the case where \( \partial_j \sigma(x) > 0 \), the case \( \partial_j \sigma(x) < 0 \) being similar. Note first that \( P(D = 1|X) > 0 \) almost surely, because \( P(D = 1|X = x) = E[P(D = 1|X = x, Y^*)|X = x] \), and \( P(D = 1|X = x, Y^* = y) \) is bounded from below by \( h/2 > 0 \) for \( y \) large enough. Now, consider the lower bound. We have for
all \((u, x)\), by independence between \(\varepsilon\) and \(X\),
\[
P(\varepsilon \leq u) = P(\varepsilon \leq u | X = x) \\
\leq P(\varepsilon \leq u, D = 1 | X = x) + P(D = 0 | X = x) \\
\leq P(Y \leq \psi(x) + \sigma(x) u | D = 1, X = x) P(D = 1 | X = x) + P(D = 0 | X = x).
\]

Taking \(u = Q_\varepsilon(\tau)\), using \(F_\varepsilon(Q_\varepsilon(\tau)) \geq \tau\) and the definition of the quantiles of \(Y | D = 1, X = x\), we obtain, for all \(x\) in the support of \(X\),
\[
\psi(x) + \sigma(x) Q_\varepsilon(\tau) \geq Q_{Y|D=1,X=x}(\frac{\tau - P(D = 0 | X = x)}{P(D = 1 | X = x)}) .
\]
As a result,
\[
Q_\varepsilon(\tau) \geq \sup_{x \in \text{Supp}(X)} \frac{Q_{Y|D=1,X=x}(\frac{\tau - P(D = 0 | X = x)}{P(D = 1 | X = x)}) - \psi(x)}{\sigma(x)} .
\]
The outer lower bound of \(\Delta_{j\tau}(x)\) follows from \(\Delta_{j\tau}(x) = \partial_j \psi(x) + \partial_j \sigma(x) Q_\varepsilon(\tau)\).

Now let us turn to the outer upper bound. Reasoning as before, we have, for all \(x\) in the support of \(X\) and \(u < Q_\varepsilon(\tau)\),
\[
\tau \geq P(\varepsilon \leq u) \geq P(Y \leq \psi(x) + \sigma(x) u, D = 1 | X = x) P(D = 1 | X = x).
\]
The definition of the quantiles of \(Y | D = 1, X = x\) then yields
\[
\psi(x) + \sigma(x) u \leq Q_{Y|D=1,X=x}(\frac{\tau}{P(D = 1 | X = x)}) .
\]
Letting \(u\) tend to \(Q_\varepsilon(\tau)\) and taking the infimum over \(x\) then yields
\[
Q_\varepsilon(\tau) \leq \inf_{x \in \text{Supp}(X)} \frac{Q_{Y|D=1,X=x}(\frac{\tau}{P(D = 1 | X = x)}) - \psi(x)}{\sigma(x)} .
\]
The upper bound follows.
B.3.3 Proof of Proposition 3.2.1

We verify Assumption 21 with $h = 1$. By Assumption 23 and because $f_C(\gamma) \to 0$, we have, as $y \to \infty$,

$$|P(D = 1|X = x, Y^* = y) - 1| = \left| P \left( F_\eta(\eta) \leq F_\eta(\phi(x)) | F_{\tilde{\eta}}(\tilde{\eta}) = F_{\tilde{\eta}} \left( \frac{\psi(x) - y}{\sigma(x)} \right) \right) - 1 \right|$$

$$\leq \left| \partial_1 C \left[ F_{\tilde{\eta}} \left( \frac{\psi(x) - y}{\sigma(x)} \right), F_{\eta}(\phi(x)) \right] - 1 \right|$$

$$\leq \sup_{\in \mathbb{R}^1} \left| \partial_1 C \left[ F_{\tilde{\eta}} \left( \frac{\psi(x) - y}{\sigma(x)} \right), v \right] - 1 \right| \to 0.$$ 

B.3.4 Proof of Lemma 3.3.1

Let $U_x(y) \equiv 1/P(Y > y|X = x)$, $V_x(y) \equiv 1/hS_x((y - x\beta)/(1 + x'\delta))$. Then from Equation (3.3.1), $U_x(y) \sim V_x(y)$. We want to show the equivalence $U_x^{-}(\tau) \sim V_x^{-}(\tau)$.

For that purpose, we suppose that there exists $\varepsilon_0 > 0$ and a sequence $(y_m)_{m \in \mathbb{N}}$ tending to infinity such that

$$V_x^{-}(y_m)/U_x^{-}(y_m) \geq 1 + \varepsilon_0,$$  \hspace{1cm} (B.3.9)

and shows that this leads to a contradiction. The reasoning is similar for the other inequality $(V_x^{-}(y_m)/U_x^{-}(y_m) \leq 1 - \varepsilon_0)$.

First, by Lemma B.4.1 in Appendix D, $S_\varepsilon$ is in the domain of attraction of Type I extreme value distribution. This implies that $V \equiv 1/S_\varepsilon$ is $\Gamma$-varying (see Resnick (1987) Proposition 0.10), i.e. $\lim_{z \to \infty} \frac{V(z + tf(z))}{V(z)} = e^t$ for some auxiliary function $f$.

Define $f_x(y) = f \left[ (y - x'\beta)/(1 + x'\delta) \right] \times (1 + x'\delta)$. Then

$$\frac{V_x(z + tf_x(z))}{V_x(z)} = \frac{V \left[ \frac{z - x'\beta}{1 + x'\delta} + tf \left( \frac{z - x'\beta}{1 + x'\delta} \right) \right]}{V \left[ \frac{z - x'\beta}{1 + x'\delta} \right]} \to e^t$$

238
as $z \to \infty$. Thus $V_x(z)$ is $\Gamma$-varying with auxiliary function $f_x$. Furthermore, $U_x(z) \sim V_x(z)$ and $z + tf_x(z) \to \infty$, which implies

$$\frac{U_x(z + tf_x(z))}{U_x(z)} = \frac{U_x(z + tf_x(z))}{V_x(z + tf_x(z))} \frac{V_x(z + tf_x(z))}{U_x(z)} \frac{V_x(z + tf_x(z))}{V_x(z)} \to e^t.$$ 

Hence, $U_x$ is also $\Gamma$-varying with the same auxiliary function. $f_x$ also satisfies (see (Resnick, 1987), Ex. 0.4.3.10)

$$\lim_{z \to \infty} \frac{f_x(z)}{z} = 0. \tag{B.3.10}$$

Combining (B.3.9) and (B.3.10), we obtain that for $m$ large enough,

$$\frac{V_x^+(y_m)}{U_x^+(y_m)} \geq 1 + \varepsilon_0 \frac{f_x(U_x^+(y_m))}{U_x^+(y_m)}.$$

Now, because $y \sim V_x(V_x^+(y))$ and $y \sim U_x(U_x^+(y))$ (see (Resnick, 1987), page 28), for any $\varepsilon_1 > 0$, there exists $m$ large enough such that

$$y_m(1 + \varepsilon_1) \geq V_x(V_x^+(y_m))$$

$$\geq V_x(U_x^+(y_m)) + \varepsilon_0 f_x(U_x^+(y_m)))$$

$$\geq (1 - \varepsilon_1)U_x(U_x^+(y_m)) + \varepsilon_0 f_x(U_x^+(y_m)))$$

$$= (1 - \varepsilon_1)^2 e^{\varepsilon_0} U_x(U_x^+(y_m))$$

$$\geq (1 - \varepsilon_1)^3 e^{\varepsilon_0} y_m.$$

Therefore, $1 \geq \frac{(1 - \varepsilon_1)^3 e^{\varepsilon_0}}{1 + \varepsilon_1}$. Letting $\varepsilon_1$ tend to zero leads to a contradiction.

**B.3.5 Proof of Theorem 3.3.1**

First let us introduce additional notations. For any $\tau$, let $\theta(\tau) = (\gamma(\tau), \beta(\tau))'$. Let us also define $\tilde{Z}_n(l) = \alpha_n(l)(\tilde{\theta}(\tau_n) - \theta(\tau_n))$, with

$$\alpha_n(l) = \frac{\sqrt{\tau_n}}{\gamma(ml\tau_n) - \gamma(l\tau_n)} = \frac{\sqrt{\tau_n}}{Q_\varepsilon(ml\tau/h) - Q_\varepsilon(l\tau/h)}$$
for some arbitrary fixed $m > 1$ and $\alpha_n \equiv \alpha_n(1)$. Let also

$$\tilde{Z}_n(l_1, \ldots, l_J) = \left( \tilde{Z}'_n(1), \tilde{Z}'_n(l_1), ..., \tilde{Z}'_n(l_J) \right)' .$$

Finally, let us define

$$G_n \equiv -\frac{\partial g_n(\delta)}{\partial \delta} = \begin{pmatrix} \hat{\gamma}(l_1 \tau_n) - \hat{\gamma}(\tau_n) \\ \vdots \\ \hat{\gamma}(l_J \tau_n) - \hat{\gamma}(\tau_n) \end{pmatrix} \otimes I_d,$$

$$\tilde{G}_n = G_n / (\gamma(m \tau_n) - \gamma(\tau_n)) \text{ and}$$

$$\tilde{B}_n = \begin{pmatrix} \hat{\beta}(l_1 \tau_n) - \hat{\beta}(\tau_n) \\ \vdots \\ \hat{\beta}(l_J \tau_n) - \hat{\beta}(\tau_n) \end{pmatrix} .$$

The main part of the proof is devoted to the asymptotic normality of $\tilde{\delta}$. The asymptotic normality of $\tilde{\beta}$ follows quite easily.

The behavior of $\tilde{\delta}$ is related to $\tilde{G}_n$ and $\tilde{Z}_n(l_1, \ldots, l_J)$. To see this, note that the first order condition of (3.3.3) writes

$$G'_n W_n G_n \tilde{\delta} = G'_n W_n \tilde{B}_n .$$

Remarking that $\delta = [G'_n W_n G_n]^{-1} G'_n W_n G_n \delta$ and $\tilde{B}_n - G_n \delta = g_n(\delta)$, we obtain

$$\sqrt{n} \tilde{\tau} (\tilde{\delta} - \delta) = \left[ \tilde{G}'_n W_n \tilde{G}_n \right]^{-1} \tilde{G}'_n W_n \left[ \alpha_n g_n(\delta) \right] .$$

Moreover, some algebra shows that

$$\alpha_n g_n(\delta) = (I_J \otimes \Delta) \Gamma \tilde{Z}_n(l_1, \ldots, l_J) .$$

We thus obtain

$$\sqrt{n} \tilde{\tau} (\tilde{\delta} - \delta) = \left[ \tilde{G}'_n W_n \tilde{G}_n \right]^{-1} \tilde{G}'_n W_n (I_J \otimes \Delta) \Gamma \tilde{Z}_n(l_1, \ldots, l_J) . \tag{B.3.11}$$
The first step of the proof shows that \( \hat{Z}_n(l_1, \cdots, l_J) \) is asymptotically normal. The proof of this part is related to the proof of Theorem 5.1 in Chernozhukov (2005b), but we have to take into account that (3.3.2) is an equivalence, not an equality as in his framework. The second step establishes that \( \tilde{G}_n \overset{p}{\rightarrow} G/\log(m) \). Both steps, combined with (B.3.11), prove the asymptotic normality of \( \hat{\delta} \). We then show in the third step the main asymptotic result on \( \hat{\beta} \). Finally, Step 4 establishes the consistency of \( \hat{\beta} \) and the fact that the rate of convergence of \( \hat{\delta} \) and \( \hat{\beta} \) can be polynomial under some additional conditions on \( f(.) \).

1. \( \hat{Z}_n(l_1, \cdots, l_J) \overset{d}{\rightarrow} \mathcal{N}(0, \log(m)^{-2}L \otimes \Omega_0) \).

We prove the result for \( \hat{Z}_n(1) \) only, the multivariate generalization being straightforward but notationally cumbersome. Similarly to Chernozhukov (2005b), Equation (9.43), \( \hat{Z}_n(1) \) minimizes

\[
\Psi_n(z, \tau_n) = W_n(\tau_n)' z + \Lambda_n(z, \tau_n),
\]

with, for any \( \tau \),

\[
W_n(\tau) = \frac{-1}{\sqrt{\tau n}} \sum_{i=1}^n (\tau - 1\{ (\bar{Y}_i - \gamma(\tau) - X_i' \beta(\tau) \leq 0) \}) \bar{X}_i \tag{B.3.12}
\]

and for any \( z = (z_1, z_2)' \in \mathbb{R} \times \mathbb{R}^d \),

\[
\Lambda_n(z, \tau) = \frac{\alpha_n}{\sqrt{\tau n}} \sum_{i=1}^n \left\{ \int_0^{\|z_1 + X_i z_2\|} 1\{ \bar{Y}_i - \gamma(\tau) - X_i' \beta(\tau) \leq s \} ds \right\} \tag{B.3.13}
\]

\[
- 1\{ \bar{Y}_i - \gamma(\tau) - X_i' \beta(\tau) \leq 0 \} ds. \]

\( \Lambda_n(z, \tau_n) \) is convex in \( z \) because the integrands are increasing in \( s \). Moreover, by Lemma B.4.4 in Appendix D, \( \Lambda_n(z, \tau_n) \rightarrow \frac{1}{2} \log(m) z' \mathcal{Q}_H z \). We shall now prove that

\[
W_n(\tau_n) \overset{d}{\rightarrow} \mathcal{N}(0, \mathcal{Q}_X). \tag{B.3.14}
\]
By applying the convexity lemma and the same arguments as in the end of the proof of Theorem 1 in Pollard (1991a), Condition (B.3.14) implies $\tilde{Z}_n(1) + \log(m)^{-1} Q_H^{-1} W_n(\tau_n) = o_P(1)$ and thus $\tilde{Z}_n(1) \rightsquigarrow N(0, \log(m)^{-2} \Omega_0)$.

To establish (B.3.14), let $M_{n,i}(\tau) = \frac{1}{\sqrt{n}} (\tau - 1 \{-\hat{Y}_i - \gamma(\tau) - X'_i \beta(\tau) \leq 0\} ) \bar{X}_i + \sqrt{\tau} \mu(\tau)$, with

$$\mu(\tau) \equiv \frac{E \left[ (\tau - 1 \{-\hat{Y} \leq \gamma(\tau) + X' \beta(\tau)\} ) \bar{X} \right]}{\tau}. $$

Then

$$W_n(\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} M_{n,i}(\tau) - \sqrt{n \tau} \mu(\tau). $$

(B.3.15)

By Lemma 9.6 of Chernozhukov (2005b), we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} M_{n,i}(\tau_n) \rightsquigarrow N(0, \Omega_X). $$

(B.3.16)

Besides,

$$\| \mu(\tau) \| = \frac{1}{\tau} \left\| E \left[ (h F_{\tilde{Z}}(\gamma(\tau)) - P(D = 1, \tilde{Z} \leq \gamma(\tau)|X)) | X \right] \right\| $$

$$= \frac{1}{\tau} \left\| E \left\{ \bar{X} \int_{-\infty}^{\gamma(\tau)} [h - P(D = 1|X, \tilde{Z} = e)] d F_{\tilde{Z}}(e) \right\} \right\| $$

$$\leq \frac{1}{\tau} E \left\{ \| \bar{X} \| \sup_{e \leq \gamma(\tau)} |h - P(D = 1|X, \tilde{Z} = e)| \right\} F_{\tilde{Z}}(\gamma(\tau)) $$

$$= \frac{1}{h} f(\gamma(\tau)).$$

By Assumption 27, $\sqrt{n \tau_n} f(\gamma(\tau_n)) = o(1)$. Combined with (B.3.15) and (B.3.16), this proves (B.3.14).

2. $\tilde{G}_n \xrightarrow{p} G/ \log(m)$ and asymptotic normality of $\hat{\delta}$.
First,
\[
\frac{\tilde{\gamma}(l_{\tau_n}) - \gamma(l_{\tau_n})}{\gamma(m_{\tau_n}) - \gamma(l_{\tau_n})} = \frac{\tilde{\gamma}(l_{\tau_n}) - \gamma(l_{\tau_n})}{\gamma(m_{\tau_n}) - \gamma(l_{\tau_n})} + \frac{\gamma(l_{\tau_n}) - \gamma(\tau_n)}{\gamma(m_{\tau_n}) - \gamma(\tau_n)} + \frac{\gamma(\tau_n) - \tilde{\gamma}(l_{\tau_n})}{\gamma(m_{\tau_n}) - \gamma(\tau_n)} \quad (B.3.17)
\]

Besides, by definition of $\alpha_n$, $\tilde{\theta}(\cdot)$ and $\tilde{Z}_n(l)$, Step 1 of the proof and because $\tau_n n \to \infty$ by Assumption 27,
\[
\frac{\tilde{\gamma}(l_{\tau_n}) - \gamma(l_{\tau_n})}{\gamma(m_{\tau_n}) - \gamma(l_{\tau_n})} = \frac{\epsilon_1' \alpha_n(\tilde{\theta}(l_{\tau_n}) - \theta(l_{\tau_n}))}{\sqrt{\tau_n n}} = \frac{\epsilon_1' \tilde{Z}_n(l)}{\tau_n n} = o_p(1),
\]

where $\epsilon_1 = (1, 0, ..., 0)'$. Similarly, the third term of (B.3.17) also tends to zero in probability. Now, by Lemma B.4.3 in Appendix D,
\[
\frac{\gamma(l_{\tau_n}) - \gamma(\tau_n)}{\gamma(m_{\tau_n}) - \gamma(\tau_n)} = \frac{Q_{\tilde{z}}(l_{\tau_n} / h) - Q_{\tilde{z}}(\tau_n / h)}{Q_{\tilde{z}}(m_{\tau_n} / h) - Q_{\tilde{z}}(\tau_n / h)} \to \frac{\log(l)}{\log(m)}.
\]

Hence,
\[
\frac{\tilde{\gamma}(m_{\tau_n}) - \gamma(\tau_n)}{\gamma(m_{\tau_n}) - \gamma(\tau_n)} \overset{p}{\to} \frac{\log(l)}{\log(m)},
\]

which in turn establishes that $\tilde{G}_n \overset{p}{\to} G / \log(m)$. Combined with Step 1 and (B.3.11), this shows that
\[
\sqrt{\tau_n n}(\tilde{\delta} - \delta) \rightsquigarrow N(0, \Omega_\delta),
\]

where $\Omega_\delta = (G'WG)^{-1}G'W(I_{ij} \otimes \Delta)\Gamma(L \otimes \Omega_0)\Gamma'(I_{ij} \otimes \Delta')WG(G'WG)^{-1}$. The optimal weighting matrix is then (see, e.g., (Wooldridge, 2002), Problem 8.5)
\[
W_\delta^* = [(I_{ij} \otimes \Delta)\Gamma(L \otimes \Omega_0)\Gamma'(I_{ij} \otimes \Delta')]^{-1},
\]

and the corresponding asymptotic variance is $\Omega_\delta^* = (G'W_\delta^*G)^{-1}$.

3. Asymptotic normality of $\hat{\beta}$.
Consider first \( \hat{\beta}_j = \hat{\beta}(l_j \tau_n) + \hat{\gamma}(l_j \tau_n) \hat{\delta} \) for \( j \in \{0, \ldots, J\} \). We have

\[
\frac{\sqrt{\tau_n n}}{\gamma(\tau_n)} (\hat{\beta}_j - \beta) = \frac{\gamma(l_j \tau_n)}{\gamma(\tau_n)} \left[ \sqrt{\frac{\tau_n n}{\gamma(l_j \tau_n)}} \left( \frac{\hat{\gamma}(l_j \tau_n) - \gamma(l_j \tau_n)}{\gamma(l_j \tau_n)} \right) \hat{\delta} + \sqrt{\tau_n n} (\hat{\delta} - \delta) \right. \\
- \left. \sqrt{\frac{\tau_n n}{\gamma(l_j \tau_n)}} \left( \frac{\hat{\beta}(l_j \tau_n) - \beta(l_j \tau_n)}{\gamma(l_j \tau_n)} \right) \right]. \tag{B.3.18}
\]

By Lemma B.4.1 in Appendix D, \( \gamma(.) \in RV_0(0) \). Thus, the first ratio on the right-hand side tends to one. We now show that the first and third term in the brackets are \( o_P(1) \). We have

\[
\sqrt{\tau_n n} \frac{\hat{\gamma}(l_j \tau_n) - \gamma(l_j \tau_n)}{\gamma(l_j \tau_n)} = \left( \frac{\gamma(m \tau_n) - \gamma(l_j \tau_n)}{\gamma(l_j \tau_n)} \right) \left( \epsilon' \alpha_n (\hat{\theta}(l_j \tau_n) - \theta(l_j \tau_n)) \right). \tag{B.3.19}
\]

The second term of the right-hand side is \( \epsilon' \hat{Z}(l_j) \) and is therefore bounded in probability uniformly over \( j \). Because \( \gamma(.) \in RV_0(0) \), the first term converges to 0. Thus, the first term in the brackets of the right-hand side of (B.3.18) is a \( o_P(1) \). The same reasoning applies to the third term in the brackets in (B.3.18).

Hence,

\[
\frac{\sqrt{\tau_n n}}{\gamma(\tau_n)} (\hat{\beta}_j - \beta) = \sqrt{\tau_n n} (\hat{\delta} - \delta) + o_P(1).
\]

Now, because \( \hat{\beta} = \sum_{j=0}^{J} \hat{\beta}_j / (J + 1) \), we obtain

\[
\frac{\sqrt{\tau_n n}}{\gamma(\tau_n)} (\hat{\beta} - \beta) \xrightarrow{\text{d}} \mathcal{N}(0, \Omega_\delta).
\]

It also follows from the fact that the left-hand side of (B.3.19) converges to 0 that \( \overline{\gamma(l_j \tau_n)} / \gamma(l_j \tau_n) \xrightarrow{p} 1 \), and in particular \( \overline{\gamma(\tau_n)}/\gamma(\tau_n) \xrightarrow{p} 1 \). This implies that \( \gamma(\tau_n) \) can be replaced by \( \overline{\gamma(\tau_n)} \) in the equation above.

4. Consistency of the bootstrap
Hereafter, all bootstrap counterparts are starred. Let \( \{I_{n,j}\}_{j=1}^{\infty} \) denote an i.i.d. sequence distributed as multinomial with parameter \( \left( \frac{1}{n}, \ldots, \frac{1}{n} \right) \), so that the bootstrap weight for individual \( i \), \( w_{n,i} \), satisfies \( w_{n,i} = \sum_{j=1}^{n} 1\{I_{n,j} = i\} \).

First, note that as in (B.3.11),

\[
\sqrt{n}(\hat{\delta}^* - \delta) = \left[ \hat{G}_n W_n \hat{G}_n \right]^{-1} \hat{G}_n W_n (I_J \otimes \Delta) \Gamma \hat{Z}_n^*(l_1, \ldots, l_J). \tag{B.3.20}
\]

We now prove that \( \hat{Z}_n^*(l_1, \ldots, l_J) \) can be linearized. As above and to avoid cumbersome notations, we consider \( \hat{Z}_n^*(1) \) only. First, observe that as in (B.3.14), \( \hat{Z}_n^*(1) \) minimizes

\[
\Psi_n^*(z, \tau_n) = W_n^*(\tau_n)'z + \Lambda_n^*(z, \tau_n).
\]

By part 2 of Lemma B.4.4,

\[
\Psi_n^*(z, \tau_n) = W_n^*(\tau_n)'z + \frac{1}{2} \log(m) z' Q_X z + o_P(1).
\]

Then by applying the same argument in the proof of theorem 1 in Pollard (1991a), we obtain

\[
\hat{Z}_n^*(1) = (\log(m)Q_H)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_{n,i}(\tau - 1\{\tilde{Y}_i - \gamma(\tau) - X_i' \beta(\tau) \leq 0\})X_i + o_P(1).
\]

Since \( E(w_{n,i}) = 1 \), we have

\[
E \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_{n,i}(\tau - 1\{\tilde{Y}_i - \gamma(\tau) - X_i' \beta(\tau) \leq 0\})X_i \right] = \sqrt{n} \mu(\tau) \to 0,
\]

which implies, in turn, that

\[
\hat{Z}_n^*(1) = (\log(m)Q_H)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_{n,i}M_{n,i}(\tau_n) + o_P(1). \tag{B.3.21}
\]
Now, reasoning exactly as in Part 2, we get that $\tilde{G}_n^* \xrightarrow{p} G/\log(m)$. Combined with Equations (B.3.20) and (B.3.21), this yields

$$\sqrt{n} \tau_n (\delta^* - \delta) = [G'WG]^{-1}G'(I \otimes \Delta)\Gamma(I_{j+1} \otimes Q_H^{-1}) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_{n,i} M_{n,i}(\tau_n) + o_P(1),$$

(B.3.22)

Besides, Part 1 implies that

$$\sqrt{n} \tau_n (\delta - \delta) = [G'WG]^{-1}G'(I \otimes \Delta)\Gamma(I_{j+1} \otimes Q_H^{-1}) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} M_{n,i}(\tau_n) + o_P(1).$$

Subtracting this from (B.3.22), we obtain

$$\sqrt{n} \tau_n (\delta^* - \delta) = [G'WG]^{-1}G'(I \otimes \Delta)\Gamma(I_{j+1} \otimes Q_H^{-1}) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (w_{n,i} - 1) M_{n,i}(\tau_n) + o_P(1).$$

Now, we use the same Poisson approximation idea as in, e.g., Chapter 3.6 of van der Vaart and Wellner (1996). Let $N_n$ be a Poisson random variable with mean $n$, independent of the data and of the $\{I_{n,j}\}_{j \geq 1}$. Let also $w_{N_n,i} = \sum_{j=1}^{N_n} 1\{I_{n,j} = i\}$, so that $\{w_{N_n,i}\}_{i=1}^{n}$ are i.i.d. Poisson random variable with unit mean. The idea is to approximate $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (w_{n,i} - 1) M_{n,i}(\tau_n)$ by $\frac{1}{\sqrt{N_n}} \sum_{i=1}^{n} (w_{N_n,i} - 1) M_{n,i}(\tau_n)$, and then apply the central limit theorem to the latter. First consider the approximation. Fix $\eta > 0$ and let $I_j = \{i : |w_{N_n,i} - w_{n,i}| \geq j\}$ and $n_j = \#I_j$. Then, for $n$ large enough and with a probability greater than $1 - \eta$ (see (van der Vaart and Wellner, 1996), p.348),

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (w_{N_n,i} - w_{n,i}) M_{n,i}(\tau_n) = \text{sign}(N_n - n) \frac{2}{\sqrt{n}} \sum_{j=1}^{2} \sqrt{n} \sum_{i \in I_j} M_{n,i}(\tau_n),$$

(B.3.23)

with the convention that $\sum_{i \in I_j} M_{n,i}(\tau_n) = 0$ when $n_j = 0$. Let us show that
\[ \sum_{i \in I_j} M_{n,i}(\tau_n) / \sqrt{n} = o_P(1). \] First, observe that

\[
E \left[ \sum_{i \in I_j} M_{n,i}(\tau_n) \big| (I_{n,j})_{j \geq 1}, N_n \right] = 0. \tag{B.3.24}
\]

Besides, because \( V(M_{n,i}(\tau_n)) \leq 1 \) for \( n \) large enough, we have, for \( n \) large enough,

\[
V \left[ \frac{1}{\sqrt{n}} \sum_{i \in I_j} M_{n,i}(\tau_n) \big| (I_{n,j})_{j \geq 1}, N_n \right] \leq \frac{n_j}{n} \leq \frac{|N_n - n|}{n}.
\]

Thus, using a decomposition of variance, Equation (B.3.24) and Jensen’s inequality, we get

\[
V \left[ \frac{1}{\sqrt{n}} \sum_{i \in I_j} M_{n,i}(\tau_n) \right] \leq \frac{V(N_n)}{n} = \frac{1}{\sqrt{n}}.
\]

This implies that \( \sum_{i \in I_j} M_{n,i}(\tau_n) / \sqrt{n} = o_P(1) \). Thus, in view of (B.3.23),

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (w_{N_n,i} - w_{n,i}) M_{n,i}(\tau_n) = o_P(1).
\]

As a result,

\[
\sqrt{n} \tau_n (\hat{\delta}^* - \hat{\delta}) = [G'WG]^{-1}G'W(I_j \otimes \Delta) \Gamma (I_{J+1} \otimes Q_H^{-1}) \frac{1}{\sqrt{N_n}} \sum_{i=1}^{n} (w_{N_n,i} - 1) M_{n,i}(\tau_n) + o_P(1).
\]

Because the \( \{w_{N_n,i} - 1\}_{i=1}^{n} \) are i.i.d., independent of the data and satisfy \( E(w_{N_n,i} - 1) = 0 \) and \( V(w_{N_n,i} - 1) = 1 \), we obtain, conditional on the data and with probability approaching one (see, e.g., Lemma 2.9.5 of (van der Vaart and Wellner, 1996)),

\[
\sqrt{n} \tau_n (\hat{\delta}^* - \hat{\delta}) \rightsquigarrow N(0, \Omega_\delta).
\]

Hence, the bootstrap is valid for \( \hat{\delta} \).
Now, a similar reasoning as in Part 3 yields

$$\frac{\sqrt{\tau_n n}}{\gamma(\tau_n)} (\hat{\beta}^*_j - \hat{\beta}) = \sqrt{\tau_n n} (\hat{\delta}^* - \hat{\delta}) + o_P(1).$$

Then the equality $$\hat{\beta}^* = \sum_{j=0}^J \hat{\beta}^*_j / (J + 1)$$ yields

$$\frac{\sqrt{\tau_n n}}{\gamma(\tau_n)} (\hat{\beta}^* - \hat{\beta}) \xrightarrow{\text{d}} \mathcal{N}(0, \Omega),$$

which show that the bootstrap is consistent for $$\hat{\beta}$$ as well. This concludes the proof of Theorem 3.3.1.

**B.3.6 Proof of Proposition 3.3.1**

First, suppose that $$f(u) = o(|u|^{-a})$$ as $$u \to -\infty$$, for some $$a > 1$$. Fix $$\alpha \in (0, 1)$$ such that $$a(1 - \alpha) > 1$$ and let $$\tau_n = \tau_n^*$$ be defined by (3.3.4). As shown in the discussion before Theorem 3.3.1, such a $$\tau_n$$ satisfies Assumption 27. Moreover, since $$\gamma(\tau_n) = G^{-1}(1/n)$$ (with $$G(\gamma) = F_{\hat{\gamma}}(\gamma) f(\gamma)^{2(1-\alpha)}$$),

$$\frac{\sqrt{\tau_n n}}{\gamma(\tau_n)} = \frac{\sqrt{h}}{f^{1-\alpha}(G^{-1}(1/n)) G^{-1}(1/n)}.$$

Because $$f^{1-\alpha}(u) u \to 0$$ as $$u \to -\infty$$ and $$G^{-1}(1/n) \to -\infty$$, we get

$$\lim_{n \to \infty} \frac{\sqrt{\tau_n n}}{\gamma(\tau_n)} = -\infty.$$

Thus, $$\hat{\beta}$$ is consistent with such a choice of $$\tau_n$$.

Now, suppose that $$f(u) = o(F_{\hat{\gamma}}(u)^a)$$ for some $$a > 0$$. Consider in this case $$\tau_n = n^{-1/(2a+1)}$$. Then $$\tau_n \to 0$$ and $$n\tau_n \to \infty$$. Because $$f(\gamma(\tau)) = o(\tau^a)$$, we also have

$$\sqrt{\tau_n n} f(\gamma(\tau_n)) = n^{a/(2a+1)} o\left(n^{-a/(2a+1)}\right) = o(1).$$

248
Hence, this choice of $\tau_n$ satisfies Assumption 27. Besides, $\gamma(.) \in RV_\alpha(0)$. This implies that for any $\alpha > 0$, $|\gamma(\tau_n)| < \tau_n^{-\alpha}$ for $n$ large enough. Choose $0 < \alpha < a$. Then, for $n$ large enough,

$$\frac{n^{(a-n)/(2a+1)}}{\gamma(\tau_n)} > \gamma(\tau_n).$$

This ensures that $\hat{\beta}$ has a polynomial rate of convergence. With such a $\tau_n$, the rate of convergence of $\hat{\delta}$ is $n^{\alpha/(2a+1)}$, which is also polynomial.

**B.3.7 Verification of (3.2.4) and (3.3.5) for several copulas**

Case 1: Gaussian copula with $\rho > 0$. We just check (3.3.5), which is stronger than (3.2.4). We have, after some algebra,

$$1 - \partial_1 C_\rho(u, v)$$

$$= 1 - \frac{1}{\varphi(\Phi^{-1}(u))} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1 - \rho^2}} \exp\left(-\frac{(\Phi^{-1}(u))^2 - 2\rho\Phi^{-1}(u)s + s^2}{2(1 - \rho^2)}\right) ds$$

$$= 1 - \frac{e^{-(1-\rho^2)\Phi^{-1}(u)^2/[2(1-\rho^2)]}}{2\pi\varphi(\Phi^{-1}(u))} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{\sqrt{(2\pi)(1-\rho^2)}} \exp\left(-\frac{(s - \rho\Phi^{-1}(u))^2}{2(1 - \rho^2)}\right) ds$$

$$= \Phi\left(\frac{\rho\Phi^{-1}(u) - \Phi^{-1}(v)}{\sqrt{1 - \rho^2}}\right).$$

Thus, because $\rho > 0$,

$$\sup_{v \in [\xi, 1], u \leq \tau} 1 - \partial_1 C_\rho(u, v) = \Phi\left(\frac{\rho\Phi^{-1}(\tau) - \Phi^{-1}(v)}{\sqrt{1 - \rho^2}}\right).$$

Now, as $x \to -\infty$, we have $\Phi(x) \sim -\varphi(x)/x$. Because for any $K > 0$, $\exp(-Kx^2) \leq -1/x \leq 1$ for $x$ small enough, we have $\varphi(x/\sigma) \leq \Phi(x) \leq \varphi(x)$ for any $0 < \sigma < 1$. This also implies that $\Phi^{-1}(\tau) \leq \sigma\varphi^{-1}(\tau)$, for $\tau$ small enough and with $\varphi^{-1}$ the inverse of $\varphi$ on $(-\infty, 0]$. Similarly, for any $m > 0$, there exists $\sigma > 1$ such that for any
x small enough, \( \varphi(x + m) \leq \varphi(x/\sigma) \). Combining these inequalities, we obtain, for any \( K < \rho/\sqrt{1 - \rho^2} \),

\[
f_C(\tau) \leq \varphi(K\varphi^{-1}(\tau)) = K' \varphi(\Phi^{-1}(\tau))^{K^2} \leq \sqrt{2\pi}^{-K^2-1} \tau^{K^2}.
\]

The result follows.

Case 2: Archimedean copulas with \( \lim_{u \to 0} \Psi(u) = +\infty \) and \( \Psi \in RV_\alpha(0) \) with \( \alpha \in (0, +\infty] \). Because \( \Psi \) is decreasing, we have, by Proposition 0.8 of Resnick (1987), \( \Psi^{-1} \in RV_{1/\alpha}(\infty) \). As a result, for all \( v \in [\underline{v}, 1] \),

\[
u \geq C(u, v) \geq \Psi^{-1}(\Psi(u) + \Psi(v)) \sim \Psi^{-1}(\Psi(u)) = u \text{ as } u \to 0.
\]

In other words,

\[
\lim_{u \to 0} \sup_{v \in [\underline{v}, 1]} |C(u, v)/u - 1| = 0.
\]

This implies that

\[
\sup_{v \in [\underline{v}, 1]} \left| \frac{\Psi'(u)}{\Psi'(C(u, v))} - 1 \right| = \left| \frac{\Psi'(u)}{\Psi'(l(u)u)} - 1 \right|. \tag{B.3.25}
\]

for some function \( l(.) \) tending to one as \( u \to 0 \). Now, by Proposition 0.7 of Resnick (1987), \( \Psi' \in RV_{\alpha-1}(0) \). This implies that the left-hand side of (B.3.25) tends to 0. (3.2.4) follows by remarking that \( \partial_1 C(u, v) = \Psi'(u)/\Psi' \circ C(u, v) \).

Case 3: Gumbel copulas with \( \theta > 1 \). Some algebra yields

\[
\partial_1 C(u, v) = \frac{1}{1 + \Psi(v; \theta)/\Psi(u; \theta)} \frac{C(u, v) \log C(u, v)}{u \log u}.
\]

Now, by the fact that \( x \log(x) \) is decreasing when \( x \) is close to 0 and \( C(u, v) \leq u \), we have \( C(u, v) \log C(u, v) \geq u \log(u) \), i.e. \( \frac{C(u, v) \log C(u, v) \leq 1}{u \log u} \leq 1 \). Because \( v \mapsto C(u, v) \) is increasing, \( C(u, v) \log C(u, v) \leq C(u, \underline{v}) \log C(u, \underline{v}) \). Furthermore, \( 0 \leq \Psi(v, \theta) \leq \)
\(\Psi(v, \theta)\). Therefore, we have

\[
\sup_{v \in [u, 1]} \left| \hat{\partial}_1 C(u, v) - 1 \right|
\leq \sup_{v \in [u, 1]} \left( \left| \frac{C(u, v) \log C(u, v)}{u \log u} - 1 \right| + \left| \hat{\partial}_1 C(u, v) - \frac{C(u, v) \log C(u, v)}{u \log u} \right| \right)
\leq \sup_{v \in [u, 1]} \left( 1 - \frac{C(u, v) \log C(u, v)}{u \log u} \right) + \sup_{v \in [u, 1]} \left( \frac{\Psi(v, \theta)}{\Psi(v, \theta) + \Psi(u, \theta)} \right) \frac{C(u, v) \log C(u, v)}{u \log u}
\leq \left( 1 - \frac{C(u, v) \log C(u, v)}{u \log u} \right) + \frac{\Psi(v, \theta)}{\Psi(v, \theta) + \Psi(u, \theta)}
\]

\(\Psi(u, \theta) \to \infty\) as \(u \to 0\), so the second term also converges 0. Therefore, to prove (3.2.4), it suffices to show that \(C(u, v) \sim u\). We have, for \(\theta > 1\),

\[
C(u, v) = \exp \left[ - \left( (- \log u)^\theta + (- \log v)^\theta \right)^{1/\theta} \right]
= \exp \left[ \log u \left( 1 + \left( \frac{- \log v}{- \log u} \right)^\theta \right)^{1/\theta} \right]
= \exp \left[ \log u + \frac{(- \log v)^\theta}{\theta(- \log u)^{\theta-1}} + o \left( \frac{1}{(- \log u)^{\theta-1}} \right) \right]
\sim u.

Case 4: Clayton copula with \(\theta > 0\). We obtain in this case

\[
1 - \hat{\partial}_1 C(u, v; \theta) \leq K u^\theta \left( \frac{1}{v^\theta} - 1 \right)
\]

Hence, \(f_C(\tau) \leq K' \tau^\theta\), where \(K' = K \left( \frac{1}{\tau^\theta} - 1 \right)\). (3.3.5) follows.

Case 5: Rotated Gumbel-Barnett copula with \(\theta \in (0, 1]\). We have

\[
1 - \hat{\partial}_1 C(u, v; \theta) = (1 - v) \exp(-\theta \log(u) \log(1-v))(1 - \theta \log(1-P)) \leq O \left( u^{-\theta \log(1-v)} \right)
\]

(B.3.26)
It follows that (3.3.5) holds.

**Case 6:** \( C(u, v; \theta) = (1 + [(u^{-1} - 1)^{\theta} + (v^{-1} - 1)^{\theta}]^{1/\theta})^{-1} \) with \( \theta > 1 \). In this case,

\[
1 - \partial_1 C(u, v; \theta) = 1 - \left( \frac{1}{u + [(1 - u)^{\theta} + u^{\theta}(v^{-1} - 1)^{\theta}]^{1/\theta}} \right)^2 \left[ 1 + \left( \frac{u^{-1} - 1}{u^{-1} - 1} \right)^{\theta} \right]^{1/\theta-1} \leq K u.
\]

(3.3.5) follows.

**Case 7:** \( C(u, v; \theta) = (1 + [(u^{-1/\theta} - 1)^{\theta} + (v^{-1/\theta} - 1)^{\theta}]^{1/\theta})^{-\theta} \) with \( \theta \geq 1 \). We have

\[
\partial_1 C(u, v; \theta) = 1 - (u^{1/\theta} + [(1 - u^{1/\theta})^{\theta} + u(v^{-1/\theta} - 1)^{\theta}]^{1/\theta})^{-1} \left[ 1 + \left( \frac{u^{\theta} - 1}{u^{\theta} - 1} \right)^{\theta} \right]^{1/\theta-1} \leq K u^{1/\theta}
\]

which implies (3.3.5).

**Case 8:** \( C(u, v; \theta) = \theta/\log(\exp(\theta/u) + \exp(\theta/v) - \exp(\theta)) \) with \( \theta > 0 \). We have

\[
1 - \partial_1 C(u, v; \theta) = 1 - 1/(1 + \log(1 + (\exp(\theta/v) - \exp(\theta)) \exp(-\theta/u)))^2 \frac{1}{1 + (\exp(\theta/v) - \exp(\theta)) \exp(-\theta/u)} \leq K \exp(-\theta/u)
\]

Thus Condition (3.3.5) is easily satisfied. In this case, any polynomial rate slower than the parametric rate is in fact possible.

**Case 9:** \( C(u, v; \theta) = [\log(\exp(u^{-\theta}) + \exp(v^{-\theta}) - e)]^{-1/\theta} \) with \( \theta > 0 \). Start from

\[
1 - \partial_1 C(u, v; \theta) = 1 - \left[ 1 + u^\theta \log \left( 1 + \frac{\exp(v^{-\theta}) - e}{\exp(u^{-\theta})} \right) \right]^{-1/\theta-1} \frac{1}{1 + \frac{\exp(v^{-\theta}) - e}{\exp(u^{-\theta})}} \leq K_1 u^\theta \log \left( 1 + \left[ \exp(v^{-\theta}) - e \right] \exp(-u^{-\theta}) \right) + K_2 \exp(-u^{-\theta}) \leq K \exp(-u^{-\theta})
\]
Therefore, Condition (3.3.5) is easily satisfied and once more, any polynomial rate slower than parametric rate is possible.

**B.3.8 Proof of Theorem 3.3.2**

We use the notations of the proof of Theorem 3.3.1 (see Subsection C.5) along with the notations introduced in the text before Theorem 3.3.2.

Reasoning as previously, we have

\[
\log(m)\alpha_n(\hat{\Delta}^1 - \Delta^1) = (G'_\Delta W_1 G_\Delta)^{-1} G'_\Delta W_1 (\Gamma_3 \otimes \Psi \Gamma_2) \log(m) \hat{Z}_n(l_1, \ldots, l_J) + o_P(1).
\]

Remark that \(\log(m)\alpha_n = \sqrt{n\tau_n} \lambda_n / \gamma(\tau_n)\), with \(\lambda_n = \gamma(\tau_n) \log(m) / [\gamma(m\tau_n) - \gamma(\tau_n)]\). Then the asymptotic normality of \(\hat{Z}_n(l_1, \ldots, l_J)\) yields

\[
\lambda_n \frac{\sqrt{\tau_n n}}{\gamma(\tau_n)} (\hat{\Delta}^1 - \Delta^1) \rightsquigarrow \mathcal{N}(0, (G'_\Delta W_1 G_\Delta)^{-1} G'_\Delta W_1 (\Gamma_3 \otimes \Psi \Gamma_2)(L \otimes \Omega_0)(\Gamma'_3 \otimes \Gamma'_2 \Psi') W_1 G_\Delta (G'_\Delta W_1 G_\Delta)^{-1}).
\]

It follows from the proof of Theorem 3.3.1 that \(\hat{\lambda}_n / \lambda_n \xrightarrow{p} 1\) and \(\gamma(\tau_n) / \gamma(\tau_n) \xrightarrow{p} 1\). Therefore, we can replace \(\lambda_n\) by \(\hat{\lambda}_n\) and \(\gamma(\tau_n)\) by \(\gamma(\tau_n)\) in the previous equation. Finally, the optimal matrix is

\[
W^*_\Delta = [(\Gamma_3 \otimes \Psi \Gamma_2)(L \otimes \Omega_0)(\Gamma'_3 \otimes \Gamma'_2 \Psi')]^{-1}
\]

and the corresponding asymptotic variance is \((G'_\Delta W^*_\Delta G_\Delta)^{-1}\). Finally, exactly the same reasoning as in the proof of Theorem 3.3.1 applies for proving the consistency of the bootstrap, so its proof is omitted.

**B.4 Technical lemmas**

**Lemma B.4.1.** If Assumption 20 (ii)-(iii) hold, then \(S_\varepsilon\) is rapidly varying at \(+\infty\), i.e. its extreme value index is 0. Moreover, \(Q_\varepsilon \in RV_\varepsilon(0)\).
Proof. Because sup(Supp(\varepsilon)) = \infty, S_{\varepsilon} is not in the attraction domain of type III extreme value distributions (see (Resnick, 1987), Proposition 1.13). Suppose S_{\varepsilon} is not rapidly varying. Then, S_{\varepsilon} is not either in the attraction domain of type I extreme value distribution (See (Resnick, 1987), Exercise 1.1.9). So S_{\varepsilon} is in the attraction domain of type II extreme value distribution, i.e. S_{\varepsilon} \in RV_{-\xi^{-1}}(+\infty) with extreme value index \xi > 0. We also have

$$\frac{S_{\exp(\varepsilon)}(tx)}{S_{\exp(\varepsilon)}(x)} = \frac{S_{\exp}(u(x) \log(x))}{S_{\exp}(\log(x))}$$  \hspace{1cm} (B.4.1)$$

where u(x) = \frac{\log(l)+\log(x)}{\log(x)} \to 1 as x \to +\infty. Because S_{\varepsilon} \in RV_{-\xi^{-1}}(+\infty), the right-hand side of Equation (B.4.1) converges to 1. This implies that S_{\exp(\varepsilon)} is slowly varying, a contradiction. Thus, S_{\varepsilon} is rapidly varying at +\infty.

To prove the second result, note that 1/S_{\varepsilon} is nondecreasing, rapidly varying at +\infty and satisfies 1/S_{\varepsilon}(+\infty) = +\infty. Thus, by Proposition 0.8 of Resnick (1987), (1/S_{\varepsilon})^{-} \in RV_{0}(\infty). Remark that (1/S_{\varepsilon})^{-}(1/\tau) = -Q_{\varepsilon}(\tau). Hence, Q_{\varepsilon} \in RV_{0}(0). \ \Box

Lemma B.4.2. Suppose that Assumptions 20 (ii)- (iii) and 26 hold. Then Q_{\varepsilon}(\varepsilon\tau) - Q_{\varepsilon}(\tau) \in RV_{0}(0), Q'_{\varepsilon} \in RV_{-1}(0) and for any positive (l, m),

$$\lim_{\tau \to 0} \frac{Q_{\varepsilon}(l\tau) - Q_{\varepsilon}(\tau)}{Q_{\varepsilon}(m\tau) - Q_{\varepsilon}(\tau)} = \frac{\log(l)}{\log(m)}.$$  \hspace{1cm} (B.4.2)$$

Proof. We first prove the last point. By Lemma B.4.1, F_{\varepsilon} is in the attraction domain of type I distribution. Then by Proposition 0.10 in Resnick (1987), \tau \mapsto -Q_{\varepsilon}(\tau) is \Pi-varying with auxiliary function \tau \mapsto Q_{\varepsilon}(\varepsilon\tau) - Q_{\varepsilon}(\tau), namely

$$\lim_{\tau \to 0} \frac{Q_{\varepsilon}(l\tau) - Q_{\varepsilon}(\tau)}{Q_{\varepsilon}(m\tau) - Q_{\varepsilon}(\tau)} = \log(l)$$

for all l > 0. Then

$$\frac{Q_{\varepsilon}(l\tau) - Q_{\varepsilon}(\tau)}{Q_{\varepsilon}(m\tau) - Q_{\varepsilon}(\tau)} = \frac{Q_{\varepsilon}(l\tau) - Q_{\varepsilon}(\tau)}{Q_{\varepsilon}(e\tau) - Q_{\varepsilon}(\tau)} \cdot \frac{Q_{\varepsilon}(m\tau) - Q_{\varepsilon}(\tau)}{Q_{\varepsilon}(e\tau) - Q_{\varepsilon}(\tau)} \to \frac{\log(l)}{\log(m)}.$$
Turning to the first point, we have, by (B.4.2),

$$\frac{Q_\varepsilon(ex\tau) - Q_\varepsilon(x\tau)}{Q_\varepsilon(e\tau) - Q_\varepsilon(\tau)} = \frac{Q_\varepsilon(ex\tau) - Q_\varepsilon(\tau)}{Q_\varepsilon(e\tau) - Q_\varepsilon(\tau)} - \frac{Q_\varepsilon(x\tau) - Q_\varepsilon(\tau)}{Q_\varepsilon(e\tau) - Q_\varepsilon(\tau)} \to \log(\varepsilon) - \log(x) = 1.$$

Finally, let us prove the second point. By monotonicity of $Q_\varepsilon$,

$$\frac{Q_\varepsilon'(b\tau)(b-a)}{Q_\varepsilon(e\tau) - Q_\varepsilon(\tau)} \geq \frac{Q_\varepsilon(b\tau) - Q_\varepsilon(a\tau)}{Q_\varepsilon(e\tau) - Q_\varepsilon(\tau)} \geq \frac{Q_\varepsilon'(a\tau)(b-a)}{Q_\varepsilon(e\tau) - Q_\varepsilon(\tau)},$$

for any $b > a > 0$. Therefore, using (B.4.2),

$$\limsup_{\tau \to 0} \frac{Q_\varepsilon'(a\tau)\tau}{Q_\varepsilon(e\tau) - Q_\varepsilon(\tau)} \leq \frac{\log(b) - \log(a)}{b - a}.$$

Letting $b \downarrow a$, we obtain

$$\limsup_{\tau \to 0} \frac{Q_\varepsilon'(a\tau)\tau}{Q_\varepsilon(e\tau) - Q_\varepsilon(\tau)} \leq \frac{1}{a},$$

for any $a > 0$. Similarly, we obtain from the other inequality

$$\liminf_{\tau \to 0} \frac{Q_\varepsilon'(b\tau)\tau}{Q_\varepsilon(e\tau) - Q_\varepsilon(\tau)} \geq \frac{1}{b},$$

for any $b > 0$. By letting $a = b = 1$, we obtain

$$Q_\varepsilon'(\tau) \sim \frac{Q_\varepsilon(e\tau) - Q_\varepsilon(\tau)}{\tau} \quad (B.4.3)$$

This, combined with $Q_\varepsilon(e\tau) - Q_\varepsilon(\tau) \in RV_0(0)$, shows the second point. □

**Lemma B.4.3.** Suppose that Assumptions 19-21, 24 and 26 hold. Then, for all $x \in \text{Supp}(X)$,

$$\lim_{\tau \to 0} \left| \frac{Q_{4X}(\tau|x) - Q_\varepsilon(\tau/h)(1 + x'\delta)}{Q_\varepsilon(m\tau) - Q_\varepsilon(\tau)} \right| = 0,$$

$$f_{4X}(Q_{4X}(\tau|x)) \sim h f_\varepsilon(Q_\varepsilon(\tau/h))/(1 + x'\delta). \quad (B.4.4)$$

255
Proof. For the first point, fix $\Delta \in (0, h)$ and remark first that by Lemma B.4.2,

\[
\lim_{\tau \to 0} \frac{Q_\varepsilon(\tau/(h + \Delta)) - Q_\varepsilon(h)}{Q_\varepsilon(m\tau) - Q_\varepsilon(\tau)} = \lim_{\tau \to 0} \left[ \frac{Q_\varepsilon(\tau/(h + \Delta)) - Q_\varepsilon(\tau)}{Q_\varepsilon(m\tau) - Q_\varepsilon(\tau)} - \frac{Q_\varepsilon(\tau/h) - Q_\varepsilon(\tau)}{Q_\varepsilon(m\tau) - Q_\varepsilon(\tau)} \right] \to -\frac{\log(h + \Delta)}{\log(m)} + \frac{\log(h)}{\log(m)} = \frac{\log[h/(h + \Delta)]}{\log(m)} \tag{B.4.5}
\]

and the same holds replacing $\Delta$ by $-\Delta$.

Besides, by definition of the quantiles of $\varepsilon|X = x$, we have, for all $\tau$ small enough,

\[
\tau \leq P(\varepsilon \leq Q_{\delta X}(\tau|x)|X = x) = P(Y \geq x'\beta - Q_{\delta X}(\tau|x)|X = x) = P(Y^* \geq x'\beta - Q_{\delta X}(\tau|x), D = 1|X = x) = \int_{x'\beta - Q_{\delta X}(\tau|x)}^\infty P(D = 1|Y^* = y, X = x) dP_{Y^*|X=x}(y).
\]

For $\tau$ small enough, $P(D = 1|Y^* = y, X = x) \in [h - \Delta, h + \Delta]$ for all $y > x'\beta - Q_{\delta X}(\tau|x)$. Thus,

\[
\tau \leq (h + \Delta) P\left[\varepsilon(1 + x'\delta) \geq Q_{\delta X}(\tau|x)\right].
\]

Similarly, using $\tau \geq (h - \Delta) P(\varepsilon < Q_{\delta X}(\tau|x)|X = x)$,

\[
\tau \geq (h - \Delta) P\left[\varepsilon(1 + x'\delta) \geq Q_{\delta X}(\tau|x)\right].
\]

Then, by definition of the quantiles of $\varepsilon$,

\[
(1 + x'\delta)Q_\varepsilon(\tau/(h + \Delta)) \leq Q_{\delta X}(\tau|x) \leq (1 + x'\delta)Q_\varepsilon(\tau/(h - \Delta)).
\]

This, together with Equation (B.4.5)

\[
\limsup_{\tau} \left| \frac{Q_{\delta X}(\tau|x) - Q_\varepsilon(\tau/h)(1 + x'\delta)}{Q_\varepsilon(m\tau) - Q_\varepsilon(\tau)} \right| \leq (1 + x'\delta) \limsup_{\tau} \frac{\max \{Q_\varepsilon(\tau/(h - \Delta)) - Q_\varepsilon(\tau/h), Q_\varepsilon(\tau/h) - Q_\varepsilon(\tau/(h + \Delta))\}}{Q_\varepsilon(m\tau) - Q_\varepsilon(\tau)} \leq (1 + x'\delta) \frac{\max \{\log(h/(h - \Delta)), \log((h + \Delta)/h)\}}{\log(m)}.
\]

256
By letting $\Delta$ tend to 0, the left-hand side tends to zero. The first result follows.

Now let us turn to the second result. We first show that for any fixed $x$, $Q_{\Delta X}(\tau|x)$ is II-varying. We have

$$\frac{Q_{\Delta X}(m\tau|x) - Q_{\Delta X}(\tau|x)}{Q_{\tilde{\epsilon}}(e\tau) - Q_{\tilde{\epsilon}}(\tau)} = \frac{Q_{\Delta X}(m\tau|x) - (1 + x'd)Q_{\tilde{\epsilon}}(m\tau/h)}{Q_{\tilde{\epsilon}}(e\tau) - Q_{\tilde{\epsilon}}(\tau)} - \frac{Q_{\Delta X}(\tau|x) - (1 + x'd)Q_{\tilde{\epsilon}}(\tau/h)}{Q_{\tilde{\epsilon}}(e\tau) - Q_{\tilde{\epsilon}}(\tau)} + (1 + x'd)Q_{\tilde{\epsilon}}(m\tau/h) - Q_{\tilde{\epsilon}}(\tau/h).$$

By Lemma B.4.2, $\tau \mapsto Q_{\tilde{\epsilon}}(e\tau) - Q_{\tilde{\epsilon}}(\tau)$ is slowly varying. Thus, by the first result of this lemma, the first and second term converge to zero. Since $Q_{\tilde{\epsilon}}(\tau)$ is II-varying, the third term converges to $(1 + x'd) \log(m)$. Therefore

$$\frac{Q_{\Delta X}(m\tau|x) - Q_{\Delta X}(\tau|x)}{Q_{\tilde{\epsilon}}(e\tau) - Q_{\tilde{\epsilon}}(\tau)} \sim (1 + x'd) \log(m).$$

Then

$$\frac{Q_{\Delta X}(m\tau|x) - Q_{\Delta X}(\tau|x)}{Q_{\Delta X}(e\tau|x) - Q_{\Delta X}(\tau|x)} = \frac{Q_{\Delta X}(m\tau|x) - Q_{\Delta X}(\tau|x)}{Q_{\tilde{\epsilon}}(e\tau) - Q_{\tilde{\epsilon}}(\tau)} \frac{Q_{\tilde{\epsilon}}(e\tau) - Q_{\tilde{\epsilon}}(\tau)}{Q_{\Delta X}(e\tau|x) - Q_{\Delta X}(\tau|x)} \to \log(m),$$

which proves that $Q_{\Delta X}(.|x)$ is II-varying. Now, remark that for $y$ small enough,

$$P(\tilde{\epsilon} \leq y|X = x) = P(\tilde{Y} + X'b \leq y|X = x)$$

$$= P(Y^* \geq -y + x'b, D = 1|X = x)$$

$$= P\left(\tilde{\epsilon} \leq \frac{y - x'b}{1 + x'd}|D = 1, X = x\right) P(D = 1|X = x).$$

This equality, combined with Assumption 26 and the fact that $X$ is bounded, ensures that the cdf of $\tilde{\epsilon}|X$ is increasing. As a result, $Q'_{\Delta X}(.|x)$ is decreasing at the lower tail and we have, by the same reasoning as in Lemma B.4.2,

$$Q_{\Delta X}(\tau|x)' \sim \tau(Q_{\Delta X}(m\tau|x) - Q_{\Delta X}(\tau|x)). \quad (B.4.6)$$
Combining Equations (B.4.3) and (B.4.6), we obtain
\[
\frac{Q_{\bar{e}}(\tau/h)'}{Q_{dX}(\tau|x)'} \sim \frac{(Q_{\bar{e}}(m\tau) - Q_{\bar{e}}(\tau))}{h(Q_{dX}(m\tau|x) - Q_{dX}(\tau|x))} \sim \frac{1}{h(1 + x'\delta)}
\]
This proves the second result of the lemma.

Lemma B.4.4. Suppose that Assumptions 19-21 and 24-27 hold and let \( \Lambda_n(z, \tau) \) be defined as in (B.3.13). Then
\[
\Lambda_n(z, \tau_n) \overset{p}{\to} \frac{1}{2} \log(m) z' Q \varepsilon z.
\]
The same result holds for its bootstrap counterpart \( \Lambda_n^*(z, \tau_n) \).

Proof. By Lemma 9.6 in Chernozhukov (2005b), the variance of \( \Lambda_n(z, \tau) \) converges to 0. Thus it suffices to prove that \( E[\Lambda_n(z, \tau_n)] \to \frac{1}{2} \log(m) z' Q_H z \). Let us define, for any \((s, t) \in \mathbb{R}^2\),
\[
m(s, t) = \begin{cases} 
1 & \text{if } 0 < s \leq t, \\
-1 & \text{if } t \leq s < 0, \\
0 & \text{otherwise.}
\end{cases}
\]
We have
\[
E[\Lambda_n(z, \tau_n)]
\]
\[
= \frac{\alpha_n}{\sqrt{\tau_n \bar{X}}} E \left[ \int_0^{(z_1 + X' z_2)/\alpha_n} 1\{\bar{Y} - \gamma(\tau_n) - X' \beta(\tau_n) \leq s\} - 1\{\bar{Y} - \gamma(\tau_n) - X' \beta(\tau_n) \leq 0\} ds \right]
\]
\[
= \frac{n}{\sqrt{\tau_n \bar{X}}} E \left[ \int_0^{z_1 + X' z_2} 1\{\bar{Y} - \gamma(\tau_n) - X' \beta(\tau_n) \leq s/\alpha_n\} - 1\{\bar{Y} - \gamma(\tau_n) - X' \beta(\tau_n) \leq 0\} ds \right]
\]
\[
= \frac{n}{\sqrt{\tau_n \bar{X}}} E \left[ \int_0^{z_1 + X' z_2} 1\{\bar{Y} - (1 + X' \delta) Q \varepsilon(\tau_n/h) \leq s/\alpha_n\} - 1\{\bar{Y} - (1 + X' \delta) Q \varepsilon(\tau_n/h) \leq 0\} ds \right]
\]
\[
= nE \left[ \int_0^{z_1 + X' z_2} \frac{f_{dX}((1 + X' \delta) Q \varepsilon(\tau_n/h) + s/\alpha_n) - F_{dX}((1 + X' \delta) Q \varepsilon(\tau_n/h))}{\sqrt{\tau_n \bar{X}}} ds \right]
\]
\[
= E \left[ \int_{-\infty}^{+\infty} m(s, z_1 + X' z_2) s \frac{f_{dX}((1 + X' \delta) Q \varepsilon(\tau_n/h) + V_s)}{\alpha_n \sqrt{\tau_n \bar{X}}} ds \right], \quad \boxed{(B.4.7)}
\]
where for each $s$, $V_s$ is a random variable satisfying $V_s \in [0, s/\alpha_n]$. Let

$$U_n(s) = m(s, z_1 + X' z_2) s \frac{n f_{\zeta X} [(1 + X' \delta) Q_\varepsilon (\tau_n/h) + V_s]}{\alpha_n \sqrt{\tau_n n}}.$$ 

We first show that

$$U_n(s) \xrightarrow{p.s.} \frac{m(s, z_1 + X' z_2) s \log(m)}{1 + X' \delta}. \quad (B.4.8)$$

Since $1/\alpha_n = o(Q_\varepsilon (m \tau_n - Q_\varepsilon (\tau_n)))$, we have $V_s = o(Q_\varepsilon (m \tau_n - Q_\varepsilon (\tau_n)))$. Moreover, by Lemma B.4.3,

$$Q_{\zeta X} (\tau_n | x) - Q_\varepsilon (\tau_n/h)(1 + x' \delta) = o(Q_\varepsilon (m \tau_n - Q_\varepsilon (\tau_n))).$$

Then, following the same argument as Chernozhukov (2005b) after his Equation (9.57),

$$f_{\zeta X} [(1 + X' \delta) Q_\varepsilon (\tau_n/h) + V_s] \to f_{\zeta X} (Q_{\zeta X} (\tau_n | x)). \quad (B.4.9)$$

Besides, by Lemma B.4.3,

$$f_{\zeta X} (Q_{\zeta X} (\tau_n | x)) \sim h f_\varepsilon (Q_\varepsilon (\tau_n/h)) / (1 + x' \delta). \quad (B.4.10)$$

Now, by definition of $\alpha_n$ and because $Q'_\varepsilon \in RV_{-1}(0)$ by Lemma B.4.2,

$$\frac{n h f_\varepsilon (Q_\varepsilon (\tau_n/h))}{\alpha_n \sqrt{\tau_n n}} \frac{h(Q_\varepsilon (m \tau_n/h) - Q_\varepsilon (\tau_n/h)) f_\varepsilon (Q_\varepsilon (\tau_n/h))}{\tau_n}$$

$$= \left[ \int_1^m Q'_\varepsilon (s \tau_n/h) \frac{ds}{Q_\varepsilon (\tau_n/h)} \right] \to \left[ \int_1^m \frac{ds}{s} \right] = \log(m), \quad (B.4.11)$$

where the second last convergence is because, by Proposition 0.5 of Resnick (1987),

$$\frac{Q'_\varepsilon (s \tau_n/h)}{Q_\varepsilon (\tau_n/h)} \to \frac{1}{s} \text{ locally uniformly.} \frac{1}{s} \text{ is bounded over } [1, m], \text{ so dominated convergence theorem can be applied. Combining (B.4.9), (B.4.10) and (B.4.11) proves that (B.4.8) holds.}$$
Next, we prove that for \( n \) large enough,

\[
|U_n(s)| \leq U(s), \quad \text{with} \quad E \left( \int_{-\infty}^{\infty} U(s) \, ds \right) < \infty. \quad (B.4.12)
\]

Together with (B.4.8), this will allow us to use the dominated convergence theorem on the right-hand side of (B.4.7). We bound \(|U_n(s)|\) for \(|s| \leq |z_1 + X'z_2|\), since \(m(s, z_1 + X'z_2) = 0\) otherwise.

First, because \(X\) is bounded, \(\sup_{x \in \text{Supp}(X)} \gamma(\tau_n) + x'\beta(\tau_n) \to -\infty\). Thus, for any \(|s| \leq |z_1 + X'z_2|\), we have, for \(n\) large enough, \(\gamma(\tau_n) + X'\beta(\tau_n) < 0\) and \(\gamma(\tau_n) + X'\beta(\tau_n) + s/\alpha_n < 0\). Hence, by definition of \(\tilde{Y}\) and \(Y^*\),

\[
\{ \tilde{Y} \in (\gamma(\tau_n) + X'\beta(\tau_n), \gamma(\tau_n) + X'\beta(\tau_n) + s/\alpha_n) \} \\
\subseteq \{-Y^* \in (\gamma(\tau_n) + X'\beta(\tau_n), \gamma(\tau_n) + X'\beta(\tau_n) + s/\alpha_n) \}.
\]

Taking conditional expectations, this implies that for any \(|s| \leq |z_1 + X'z_2|\) and \(n\) large enough,

\[
\frac{|s|}{\alpha_n} f_{\tilde{Y}|X} [(1 + X'\delta)Q_{\tilde{Y}}(\tau_n/h) + V_s] \leq \left| F_{\tilde{Y}} \left( Q_{\tilde{Y}}(\tau_n/h) + \frac{s}{\alpha_n(1 + X'\delta)} \right) - F_{\tilde{Y}} (Q_{\tilde{Y}}(\tau_n/h)) \right|.
\]

By the mean value theorem,

\[
\left| F_{\tilde{Y}} \left( Q_{\tilde{Y}}(\tau_n/h) + \frac{s}{\alpha_n(1 + X'\delta)} \right) - F_{\tilde{Y}} (Q_{\tilde{Y}}(\tau_n/h)) \right| = \left| s \frac{F_{\tilde{Y}} (Q_{\tilde{Y}}(\tau_n/h) + V_s') - F_{\tilde{Y}} (Q_{\tilde{Y}}(\tau_n/h))}{\alpha_n(1 + X'\delta)} \right|, \quad (B.4.13)
\]

where \(V_s' \in [0, s/(\alpha_n(1 + X'\delta))]\). Because \(s/(1 + x'\delta)\) is bounded for all \(|s| \leq |z_1 + X'z_2|\) and all \(x \in \text{Supp}(X)\), \(|V_s'| \leq K/\alpha_n\). Now, by Lemma B.4.2, we have, for any \(\eta > 0\),

\[
\alpha_n [Q_{\tilde{Y}}((1 + \eta)\tau_n/h) - Q_{\tilde{Y}}(\tau_n/h)] = \sqrt{\tau_n} \frac{Q_{\tilde{Y}}((1 + \eta)\tau_n/h) - Q_{\tilde{Y}}(\tau_n/h)}{Q_{\tilde{Y}}(\tau_n/h)} \frac{Q_{\tilde{Y}}(m\tau_n/h) - Q_{\tilde{Y}}(\tau_n/h)}{Q_{\tilde{Y}}(m\tau_n/h)} \approx \sqrt{\tau_n} \frac{\log(1 + \eta)}{\log(m)} \to \infty.
\]
Hence, for $n$ large enough,

$$Q_{\bar{z}}((1 + \eta)\tau_n/h) \geq Q_{\bar{z}}(\tau_n/h) + \frac{K}{\alpha_n} \geq Q_{\bar{z}}(\tau_n/h) + V_s'.$$

Plugging this inequality in (B.4.13) and using monotonicity of $f_{\bar{z}}$, we obtain

$$\frac{|s|}{\alpha_n} f_{\bar{z}} X \left[ (1 + X' \delta) Q_{\bar{z}}(\tau_n/h) + V_s \right] \leq f_{\bar{z}} (Q_{\bar{z}}((1 + \eta)\tau_n/h)).$$

Because $1 + X'\delta$ is bounded from below, we finally get

$$U_n(s) \leq K|s| \{ |s| \leq |z_1 + X'z_2| \} \frac{n f_{\bar{z}} (Q_{\bar{z}}((1 + \eta)\tau_n/h))}{\alpha_n \sqrt{\tau_n n}}.$$

We have shown in (B.4.11) that the sequence $n f_{\bar{z}} (Q_{\bar{z}}(\tau_n/h)) / (\alpha_n \sqrt{\tau_n n})$ admits a finite limit. It is therefore bounded. Hence, we finally have, for $n$ large enough, $|U_n(s)| \leq U(s)$ with $U(s) = K|s| \{ |s| \leq |z_1 + X'z_2| \}$. Thus, (B.4.12) holds and by the dominated convergence theorem applied to (B.4.7).

$$E[G_n(z, \tau_n)] \to E \left[ \frac{\log(m)}{1 + X'\delta} \int_0^{(z_1 + X'z_2)} sds \right] = \frac{1}{2} \log(m) z' Q_H z.$$

Finally, we turn to the bootstrap counterpart $\Lambda^*_n(z, \tau)$ of $\Lambda_n(z, \tau)$, which satisfies, for all $z = (z_1, z_2)' \in \mathbb{R} \times \mathbb{R}^d$,

$$\Lambda^*_n(z, \tau) = \frac{\alpha_n}{\sqrt{\tau_n n}} \sum_{i=1}^n w_n,i \int_0^{(z_1 + X'_i z_2) / \alpha_n} \{ \bar{Y}_i - \gamma(\tau) - X'_i \beta(\tau) \leq s \} - \{ \bar{Y}_i - \gamma(\tau) - X'_i \beta(\tau) \leq 0 \} ds.$$
First,

\[ E [\Lambda_n^*(z, \tau)] \]

\[ = E [E (\Lambda_n^*(z, \tau) | \{ w_{n,i} \}_{i=1}^n)] \]

\[ = E \left\{ \sum_{i=1}^n w_{n,i} E \left[ \frac{\alpha_n}{\sqrt{\tau n}} \int_0^{(z_1 + X_i^2 z_2)/\alpha_n} \mathbb{1}\{ \tilde{Y}_i - \gamma(\tau) - X_i' \beta(\tau) \leq s \} \right] \right\} \]

\[ - \frac{1}{n} E \left\{ \sum_{i=1}^n w_{n,i} E [\Lambda_n(z, \tau)] \right\} \]

\[ = E [\Lambda_n(z, \tau)] - \frac{1}{2} \log(m) z Q_H^r z. \]

Next, we show that \( V (\Lambda_n^*(z, \tau)) \to 0 \). To see this, we note that by variance decomposition,

\[ V (\Lambda_n^*(z, \tau)) \]

\[ = E [V (\Lambda_n^*(z, \tau) | \{ w_{n,i} \}_{i=1}^n)] + V [E (\Lambda_n^*(z, \tau) | \{ w_{n,i} \}_{i=1}^n)] \]

\[ = E \left[ \sum_{i=1}^n w_{n,i}^2 \frac{\alpha_n}{\tau n} \mathbb{1} \left( \frac{(z_1 + X_i z_2^2)}{\alpha_n} \right) \mathbb{1}\{ \tilde{Y}_i - \gamma(\tau) - X_i' \beta(\tau) \leq s \} \right] \]

\[ + V \left[ \frac{1}{n} \sum_{i=1}^n w_{n,i} E (\Lambda_n(z, \tau)) \right] \]

\[ = (1 - \frac{1}{n}) V (\Lambda_n(z, \tau)) + 0 \to 0. \]

\[ \square \]
Appendix C

Appendix for Chapter 4

C.0.1 Notations

Throughout the Appendix, we denote $C$ as a generic positive constant whose value differs in different context. $L_n$ is a generic function of $n$, which is slowly varying as $n \to \infty$, i.e. $\frac{L_{kn}}{L_n} \to 1$ as $n \to \infty$ for any $k > 0$.

C.0.2 Proof for Theorem 4.3.1

For part (1), we first decompose $\hat{\alpha}_n := \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)} \hat{I}_{n,i}$ as follows:

\[
\hat{\alpha}_n = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)} \hat{I}_{n,i} \\
= \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)} I_{n,i} + \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)} \frac{f(V_i) - \hat{f}(V_i)}{f(V_i)} I_{n,i} + R_{n,1} + R_{n,2} + R_{n,3},
\]

in which

\[
R_{n,1} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)} \right) (\hat{I}_{n,i} - I_{n,i}),
\]

263
\[ R_{n,2} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{Y_i - 1\{V_i > 0\}}{f(V_i)} \right) \left( \frac{f(V_i) - \hat{f}(V_i)}{\hat{f}(V_i)} \right) (\hat{I}_{n,i} - I_{n,i}), \]

and

\[ R_{n,3} = \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{Y_i - 1\{V_i > 0\}}{f(V_i)} \right] \left[ \frac{(f(V_i) - \hat{f}(V_i))^2}{f(V_i) \hat{f}(V_i)} \right] I_{n,i}. \]

By Lemma C.0.5, the remainder terms are all asymptotically negligible, i.e.

\[ R_{n,1} + R_{n,2} + R_{n,3} = o_p\left(\frac{1}{\sqrt{n}}\right). \]

Hence, following (C.0.1), we have

\[ \hat{\alpha}_n = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i - 1\{V_i > 0\}}{f(V_i)} I_{n,i} + \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i - 1\{V_i > 0\} f(V_i) - \hat{f}(V_i)}{f(V_i)} I_{n,i} + o_p\left(\frac{1}{\sqrt{n}}\right) \]

\[ := \tilde{\delta}_{n,1} + \tilde{\delta}_{n,2} + o_p\left(\frac{1}{\sqrt{n}}\right). \] (C.0.2)

In (C.0.2), \( \tilde{\delta}_{n,2} \) represents the first-order error of the first stage kernel density estimation. Next we consider the U-decomposition of \( \tilde{\delta}_{n,2} \). Note that

\[ \tilde{\delta}_{n,2} = (C_n^2)^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} P_n(W_i, W_j), \]

in which

\[ P_n(W_i, W_j) = \frac{1}{2} \left[ \frac{Y_i - 1\{V_i > 0\}}{f^2(V_i)} \left( f(V_i) - \frac{1}{h} K\left( \frac{V_i - V_j}{h} \right) \right) I_{n,i} \right. \]

\[ + \left. \frac{Y_j - 1\{V_j > 0\}}{f^2(V_j)} \left( f(V_j) - \frac{1}{h} K\left( \frac{V_j - V_i}{h} \right) \right) I_{n,j} \right]. \]

We first claim that

\[ E(P_n^2(W_i, W_j)) = o(n). \] (C.0.3)
To see this, first recall that by Lemma C.0.6(2), on $V_i \in S_n \subset S_n^+$,
\[ f(V_i) \geq Cn^{-\rho}L_n, \]
in which $\rho = \min(\rho_r(1 + \zeta_r), \rho_l(1 + \zeta_l))$. Second, by Lemma C.0.8(2) or C.0.9(2),
\[ E\left| Y_i - 1\{V > 0\} \right|^{2+\sigma} = O(1). \]
Last, by Assumption 29(3), $f(v + hu) \leq cf(v)^{1-\sigma}$. Combining the above three facts, we have
\[
E \left[ \frac{Y_i - 1\{V_i > 0\}}{f^2(V_i)} \left( f(V_i) - \frac{1}{h}K(V_i - V_i)I_{n,i} \right) \right]^2 \\
= E \left[ \frac{Y_i - 1\{V_i > 0\}}{f^2(V_i)} \left( f^2(V_i) - 2(f(V_i) \int K(u) f(V_i + hu)du + \int \frac{K^2(u)}{h} f(V_i + hu)du)I_{n,i} \right) \right]
\leq C E \left[ \frac{Y_i - 1\{V_i > 0\}}{f^{3+\sigma}(V_i)} \right]^{2+\sigma} \frac{1}{h} I_{n,i}
= C E \left[ \frac{Y_i - 1\{V_i > 0\}}{f(V_i)} \right]^{2+\sigma} n^\rho \frac{L_n}{h}
= O\left( \frac{nL_n}{n^{1-\rho}h} \right)
= o(n).

Define
\[ \hat{U}_n = \theta_n + \frac{2}{n} \sum_{i=1}^{n} (r_n(W_i) - \theta_n), \]
in which $r_n(W_i) = E(P_n(W_i, W_j)|W_i)$, $\theta_n = Er_n(W_i)$, and $W_i = (Y_i, V_i)$. Because (C.0.3) holds, we can apply lemma 3.1 in Powell et al. (1989) and obtain that
\[ \sqrt{n}(\delta_{2,n} - \hat{U}_n) = o_p(1). \quad (C.0.4) \]
Next, we compute $r_n(W_i)$. Note that
\[ r_n(W_i) = \frac{1}{2}(r_{n,1}(W_i) + r_{n,2}(W_i)), \]
in which
\[ r_{n,1}(W_i) = E\left( \frac{Y_i - 1\{V_i > 0\}}{f^2(V_i)} (f(V_i) - \frac{1}{h} K(\frac{V_i - V_j}{h})) I_{n,i}|W_i \right) \]

and
\[ r_{n,2}(W_i) = E\left( \frac{Y_j - 1\{V_j > 0\}}{f^2(V_j)} (f(V_j) - \frac{1}{h} K(\frac{V_i - V_j}{h})) I_{n,j}|W_j \right). \]

By the mean-value theorem and Assumption 29(2), we have
\[ r_{n,1}(W_i) = \frac{Y_i - 1\{V_i > 0\}}{f^2(V_i)} I_{n,i} f^{(v)}(V_i) h^v. \]

Thus, by Lemma C.0.8(2) or Lemma C.0.9(2),
\[ E r_{n,1}^2(W_i) \leq CE\left( \frac{Y_i - 1\{V_i > 0\}}{f(V_i)} \right)^2 + \sigma I_{n,i} n^{(2-\delta)\rho} L_n h^{2v} = O(n^{(2-\sigma)\rho} h^{2v} L_n) = o(1), \]

in which the last equality is because by Assumption 33, we have \((2-\sigma)\rho < 1 < 2\nu H\).

This implies
\[ \frac{1}{n} \sum_{i=1}^{n} (r_{n,1}(W_i) - \theta_n) = o_p\left( \frac{1}{\sqrt{n}} \right). \]  
(C.0.5)

Now, we define \( P(v) = E(Y_i|V_i = v) \). For \( r_{n,2}(W_i) \), we have
\[ r_{n,2}(W_i) = E\left( \frac{Y_j - 1\{V_j > 0\}}{f(V_j)} I_{n,j} - G * K_h(V_i) \right) = E\left( \frac{Y_j - 1\{V_j > 0\}}{f(V_j)} I_{n,j} - G(V_i) - T_{n,i} \right), \]
in which
\[ G(v) = \frac{P(v) - 1\{v > 0\}}{f(v)} 1\{v \in S_n\}, \]
\[ K_h(v) = \frac{1}{h} K(\frac{v}{h}), \]
and
\[ T_{n,i} = G * K_h(V_i) - G(V_i). \]
Then
\[ r_{n,2}(W_i) - \theta_n = - \left[ \frac{P(V_i) - 1\{V_i > 0\}}{f(V_i)} \mathbb{1}\{V_i \in S_n\} - E \left( \frac{P(V_i) - 1\{V_i > 0\}}{f(V_i)} \mathbb{1}\{V_i \in S_n\} \right) \right] \]
\[ - (T_{n,i} - E(T_{n,i})). \]

If \( ET_{n,i}^2 = o(1) \), then we have
\[
\frac{1}{n} \sum_{i=1}^{n} (r_{n,2}(W_i) - \theta_n)
\]
\[= - \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{P(V_i) - 1\{V_i > 0\}}{f(V_i)} I_{n,i} - E \left( \frac{P(V_i) - 1\{V_i > 0\}}{f(V_i)} I_{n,i} \right) \right] + o_p\left( \frac{1}{\sqrt{n}} \right). \tag{C.0.6} \]

Next, we compute \( ET_{n,i}^2 \). By Minkowski Inequality, we have
\[
ET_{n,i}^2 := \int \left( \int (G(v - hu) - G(v))^2 f(v) dv \right) K(u) du
\]
\[\leq \left( \int (G(v - hu) - G(V))^2 f(v) dv \right)^{1/2} \frac{1}{2} K(u) du)^2. \tag{C.0.7} \]

By Lemma C.0.10, we have, for each fixed \( u \),
\[
\int (G(v - hu) - G(v))^2 f(v) dv \to 0
\]
as \( h \to 0 \) and
\[
\int (G(v - hu) - G(v))^2 f(v) dv \leq \int (G^2(v - hu) + G^2(v)) f(v) dv \leq C
\]
where \( C \) is a positive constant independent of \( u \). Therefore, by Dominated Convergence Theorem, the RHS of (C.0.7) vanishes as \( h \to 0 \); that is,
\[
\left[ \int \left( \int (G(V - hu) - G(V))^2 f(V) dV \right)^{1/2} K(u) du \right]^2 \to 0.
\]
This concludes the claim that $ET_{n,i}^2 = o(1)$.

Last, we note that $\theta_n = o(\frac{1}{\sqrt{n}})$ because

$$\sqrt{n}|\theta_n| \leq \sqrt{n} E|r_{n,1}(W_i)| \leq C \sqrt{n} h^v E(Y - \frac{1}{f(V)} 1\{V > 0\})^2 I_{n,i} = O((nh^{2v})^{\frac{3}{2}}) = o(1).$$

(C.0.8)

Combining (C.0.2), (C.0.5), (C.0.6), and (C.0.8), we have

$$\hat{U}_n = \theta_n + \frac{2}{n} \sum_{i=1}^{n} (r_n(W_i) - \theta_n) = \frac{1}{n} \sum_{i=1}^{n} \frac{P(V_i) - 1\{V_i > 0\}}{f(V_i)} I_{n,i} + o_p\left(\frac{1}{\sqrt{n}}\right)$$

and thus

$$\sqrt{n}(\hat{\alpha}_n - \alpha) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Y_i - P_i}{f(V_i)} I_{n,i} + o_p(1).$$

(C.0.9)

In addition, we notice that $E\left[\frac{Y_i - P_i}{f(V_i)} (1 - I_{n,i})\right]^2 \to 0$. This implies

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Y_i - P_i}{f(V_i)} I_{n,i} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Y_i - P_i}{f(V_i)} + o_p(1).$$

(C.0.10)

Combining (C.0.9) and (C.0.10), we obtain

$$\sqrt{n}(\hat{\alpha}_n - \alpha) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Y_i - P_i}{f(V_i)} + o_p(1).$$
For part (2), we first define

\[ T_{n,1} = \frac{1}{n} \sum_{i=1}^{n} 2 \left( \frac{Y_i - \mathbb{1}\{V_i > 0\}}{\hat{f}(V_i)} \right)^2 \left( \hat{I}_{n,i} - I_{n,i} \right) I_{n,i}. \]

\[ T_{n,2} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{Y_i - \mathbb{1}\{V_i > 0\}}{\hat{f}(V_i)} \right)^2 \left( \frac{f(V_i) - \hat{f}(V_i)}{\hat{f}(V_i)} \right) \hat{I}_{n,i}, \]

\[ T_{n,3} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)} \right)^2 \left( \hat{I}_{n,i} - I_{n,i} \right)^2, \]

\[ T_{n,4} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)} \right)^2 \left( \frac{f(V_i) - \hat{f}(V_i)}{\hat{f}(V_i)} \right)^2 \hat{I}_{n,i}. \]

Then

\[ \frac{1}{n} \sum_{i=1}^{n} \left( \frac{Y_i - \mathbb{1}\{V_i > 0\}}{\hat{f}(V_i)} \right)^2 \hat{I}_{n,i} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)} \right)^2 I_{n,i} + \sum_{j=1}^{4} T_{n,j}. \]

Next, we aim to show that \( T_{n,j} = o_p(1) \) for \( j = 1, \ldots, 4 \). First, since

\[ \frac{1}{n} \sum_{i=1}^{n} \left( \frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)} \right)^2 I_{n,i} \xrightarrow{p} E \left( \frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)} \right)^2 < \infty, \]

we have

\[ E|T_{n,1}| \to 0 \text{ and } T_{n,1} = o_p(1). \]

Second,

\[ |T_{n,2}| \leq \frac{1}{n} \sum_{i=1}^{n} \left( \frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)} \right)^2 o_p(1) = o_p(1). \]

Similarly, \( T_{n,3} \) and \( T_{n,4} \) can be shown to be asymptotically negligible, i.e.

\[ T_{n,3} = o_p(1), \ T_{n,4} = o_p(1). \]

Therefore, we have

\[ \frac{1}{n} \sum_{i=1}^{n} Z_{n,i}^2 \xrightarrow{p} \ E \left( \frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)} \right)^2. \]
In addition, \( \frac{1}{n} \sum_{i=1}^{n} Z_{n,i} \xrightarrow{p} \alpha \). So we have shown that

\[
\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} Z_{n,i}^2 - \left( \frac{1}{n} \sum_{i=1}^{n} Z_{n,i} \right)^2 \xrightarrow{p} \Sigma.
\]

Part (3) is just a combination of part (1) and (2).

**C.0.3 Proof of Corollary 4.3.1**

Denote \( \phi \) and \( f \) the density of \( \varepsilon \) and \( V \) with a dominating measure \( \mu \), respectively. The model (d.g.p.) \( P_\lambda \) is indexed by parameters \( \lambda = (\alpha, \phi^{\frac{1}{2}}, f^{\frac{1}{2}}) \). The collection of \( \lambda \) is \( \Lambda \subset \mathcal{H} \) where \( \mathcal{H} \) is a Hilbert space with inner product

\[
\left\langle (\alpha, \phi^{\frac{1}{2}}, f^{\frac{1}{2}}), (\alpha', \phi'^{\frac{1}{2}}, f'^{\frac{1}{2}}) \right\rangle = \alpha \alpha' + \left\langle \phi^{\frac{1}{2}} \phi'^{\frac{1}{2}} \right\rangle_{\mathcal{L}^2(\mu)} + \left\langle f^{\frac{1}{2}} f'^{\frac{1}{2}} \right\rangle_{\mathcal{L}^2(\mu)}
\]

and some dominating measure \( \mu^1 \).

Define a functional \( \psi \) that maps the model \( P_\lambda \) into \( \alpha \), i.e. \( \alpha = \psi(P_\lambda) \). Based on Lemma 25.23 in Van der Vaart (1998), in order to prove the corollary, we need to verify three conditions: (1) \( \psi(P_\lambda) \) is differentiable at \( P_\lambda \) relative to the tangent cone \( \mathcal{T}_{P_\lambda} \) in which \( P_\lambda \) satisfies the tail restrictions; (2) \( \bar{\psi} \) is the efficient score; (3)

\[
\sqrt{n}(T_n - \psi(P)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \bar{\psi}_i + o_p(1).
\]

Among them, (3) has been proved in Theorem 4.3.1. Next, we focus on (1) and (2).

We consider a one-parameter submodel \( \lambda_t = (\alpha + th, \phi_t^{\frac{1}{2}}, f_t^{\frac{1}{2}}) \) and characterize the tangent sets for \( \alpha, \phi \), and \( f \) as follows. First, \( \mathcal{T}_{P_{\lambda_t}} \), the tangent set for \( \alpha \), is \( \{ h \in \mathbb{R} \} \). To characterize the tangent set for \( \phi \), we first note that \( \eta \), the score of the submodel \( t \to \phi_t \), is defined to satisfy the following equation:

\[
\lim_{t \to 0} \int \left[ \frac{\phi_t^{\frac{1}{2}}}{t} - \phi^{\frac{1}{2}} - \frac{1}{2} \eta \phi^{\frac{1}{2}} \right]^2 dm = 0.
\]

\(^1\) Since \( \varepsilon \) and \( V \) are both assumed to be continuous random variables, \( \mu \) is just the Lebesgue measure.
Since $E\varepsilon = 0$, the submodel should also satisfy $\int \varepsilon \phi_t(\varepsilon) d\varepsilon = 0$. Then

$$E\varepsilon \eta(\varepsilon) = \langle \varepsilon, \eta(\varepsilon) \rangle_{L^2(\phi)}$$

$$= \langle 2\varepsilon \phi^{\frac{1}{2}}, \frac{1}{2} \eta \phi^{\frac{1}{2}} \rangle_{L^2(d\mu)}$$

$$= \lim_{t \to 0} \langle \varepsilon(\phi_t^{\frac{1}{2}} + \phi^{\frac{1}{2}}) - \frac{\phi_t^{\frac{1}{2}} - \phi^{\frac{1}{2}}}{t} \rangle_{L^2(d\mu)}$$

$$= \lim_{t \to 0} \int \varepsilon \phi_t(\varepsilon) d\varepsilon - \int \varepsilon \phi(\varepsilon) d\varepsilon$$

$$= 0.$$

Similarly, we can show that $E\eta(\varepsilon) = 0$. Thus $\phi \hat{P}_{\lambda t}$, the tangent set for $\phi$ is

$$\{ \eta \in L^2(\phi) : E\eta(\varepsilon) = 0, E\varepsilon \eta(\varepsilon) = 0 \}.$$

It is worthwhile to note that the tail restrictions cannot affect the tangent set. To see this, note that any $\eta$ who is continuous on a compact support and satisfies $E\eta(\varepsilon) = 0$ as well as $E\varepsilon \eta(\varepsilon) = 0$ is the score function for the submodel $t \to \phi_t = (1 + t\eta(\varepsilon)) \phi(\varepsilon)$. Because $\eta(\varepsilon)$ has compact support, $\phi_t$ has the same tail behavior as $\phi$. So the submodel satisfies the additional tail restrictions. In addition, continuous functions with compact support are dense in $L^2(\phi)$, so the (closure of) tangent set is indeed $\phi \hat{P}_{\lambda t}$.

Similarly, let $g(v)$ denote the score for $f_t$. Then the tangent set $f \hat{P}_{\lambda t}$ of $f$ is

$$\{ g \in L^2(f) : E g(V) = 0 \}.$$

We equip $\alpha \hat{P}_{\lambda t} \times \phi \hat{P}_{\lambda t} \times f \hat{P}_{\lambda t}$ with inner product

$$\langle (h, \eta, g), (h', \eta', g') \rangle = hh' + \langle \eta, \eta' \rangle_{L^2(\phi)} + \langle g, g' \rangle_{L^2(\phi^t)}. \quad (2)$$

$L^2(\phi)$ means $L^2$ norm w.r.t. probability $\phi$. 

271
Let $A: \dot{P}_{X_t} \times \phi \dot{P}_{P_{X_t}} \times f \dot{P}_{P_{X_t}} \rightarrow \mathcal{L}^2(P_{X})$ be the score operator that maps the score of $P_{X_t}$ (functions of $(V, \varepsilon)$) to the score of $P_{X_t}$ (the function of $(Y, V)$). Let $\Phi$ be the c.d.f of $\varepsilon$. Model $P_{X_t}$ has log likelihood

$$y \log(\Phi(\alpha + v)) + (1 - y) \log(1 - \Phi(\alpha + v)) + \log(f(v)). \tag{C.0.11}$$

Then by taking the ordinary derivatives of the log likelihood in (C.0.11) w.r.t. $t$, we obtain that

$$A(h, \eta, g) = \dot{i}_a h + \dot{i}_\phi \eta + \dot{i}_f g$$

where

$$\dot{i}_a = Y \frac{\phi(V + \alpha)}{\Phi(V + \alpha)} - (1 - Y) \frac{\phi(V + \alpha)}{1 - \Phi(V + \alpha)}$$

and

$$\dot{i}_f g(Y, V) = E(\varepsilon|Y, V),$$

Let $A^* : \mathcal{L}^2(P_{X}) \rightarrow \dot{P}_{X_t} \times \phi \dot{P}_{X_t} \times f \dot{P}_{X_t}$ be the adjoint of $A$. Then by the definition of adjoint, for any $b \in \mathcal{L}^2(P_{X})$,

$$\langle (h, \eta, g), A^* b \rangle := \langle A(h, \eta, g), b \rangle = \langle \dot{i}_a h + \dot{i}_\phi \eta + \dot{i}_f g, b \rangle$$

$$= \langle \dot{i}_a h, b \rangle + \langle \dot{i}_\phi \eta, b \rangle + \langle \dot{i}_f g, b \rangle$$

$$= \langle h, \langle \dot{i}_a, b \rangle_{\mathcal{L}^2(P_{X})} \rangle + \langle \eta, \dot{i}_\phi^* b \rangle + \langle g, \dot{i}_f^* b \rangle$$

$$= \langle (h, \eta, g), (\langle \dot{i}_a, b \rangle_{\mathcal{L}^2(P_{X})}, \dot{i}_\phi^* b, \dot{i}_f^* b) \rangle. \tag{C.0.12}$$

Therefore,

$$A^* b = (\langle \dot{i}_a, b \rangle_{\mathcal{L}^2(P_{X})}, \dot{i}_\phi^* b, \dot{i}_f^* b).$$

By lemma 25.34 in Van der Vaart (1998) with $X = (Y, V)$ and $Y = (V, \varepsilon)$, we have

$$\dot{i}_\phi^* b(V, \varepsilon) = E(b(Y, V)|V, \varepsilon)$$

272
and

\[ \hat{b}_f b(V, \varepsilon) = E(b(Y, V)|V, \varepsilon). \]

By Theorem 25.31 of Van der Vaart (1998), \( \psi(P_\lambda) \) is differentiable at \( P_\lambda \) relative to \( \hat{P}_{P_\lambda} \) if and only if there exists \( \tilde{\psi} \in \mathcal{L}^2(P_\lambda) \) such that \( \langle A^* \tilde{\psi}, (h, \eta, g) \rangle = h. \)

Such \( \tilde{\psi} \) is called the efficient score.

We claim that \( \tilde{\psi} = \frac{Y - E(Y|V)}{f(V)} \) satisfies all the requirement above. First, Theorem 4.3.1 has shown that, under the tail restrictions, \( \tilde{\psi} \in \mathcal{L}^2(P_\lambda) \).

In addition,

\[ \langle \tilde{\psi}, i_\alpha \rangle_{\mathcal{L}^2(P_\lambda)} = 1, \]

\[ \langle i_{\tilde{\psi}}^{\#}, \eta \rangle_{\mathcal{L}^2(P_\lambda)} = E[E(\tilde{\psi}|\varepsilon)\eta(\varepsilon)] = E\varepsilon \eta(\varepsilon) = 0, \]

and

\[ \langle i_{\tilde{\psi}}^{\#}, \eta \rangle_{\mathcal{L}^2(P_\lambda)} = E[E(\tilde{\psi}|V)g(V)] = 0. \]

This concludes that \( \psi(P_\lambda) \) is differentiable at \( P_\lambda \) relative to \( \hat{P}_{P_\lambda} \) and \( \tilde{\psi} \) is the efficient score; that is, (1) and (2) hold.

C.0.4 Proof for Theorem 4.3.2

Similar to the proof for Theorem 4.3.1, under Assumption 34(1) or (3), for any \( q_r > 0, \)

\[ \frac{C + (1 - F)^{\alpha}(z)}{(1 - F)^{\alpha}(z^{\rho})} \to 0. \]  

(C.0.13)

3 This is because we can define a functional \( \chi \) as \( \chi(\lambda_t) := \psi(P_{\lambda_t}) = \alpha + th. \) Then taking the ordinary derivative of \( \chi(\lambda_t) \) w.r.t \( t \), we have

\[ \partial_t \chi(\lambda_t) = \langle (1, 0, 0), (h, \eta, g) \rangle = h. \]
Next, we consider the integrability of variance at $+\infty$. With a change of variable,

$$
\int_0^{+\infty} \frac{1 - F_\varepsilon(\alpha + v)}{f(v)} dv = \int_0^\infty \frac{1 - F_\varepsilon(\alpha + (1 - F)^{q_r}(z))}{f((1 - F)^{q_r})(z)^2} dz
$$

$$
\geq \int_0^\infty \frac{1 - F_\varepsilon(\alpha + (1 - F)^{q_r}(z^{q_r}))}{f((1 - F)^{q_r})(z)^2} dz \tag{C.0.14}
$$

$$
= \int_0^\infty z^{q_r-2(q_r+1)} L(z) dz.
$$

Since $q_r$ can be choose to be arbitrarily small, the RHS integral will diverge at 0, which means the variance is $\infty$.

Under Assumption 34(2), there exists $q_r$ such that $q_r > \frac{\zeta}{\lambda_t}$ and $q_r - 2(\zeta + 1) \leq -1$. These two inequalities imply that (C.0.13) holds and (C.0.14) diverges to $\infty$, respectively. This concludes that $E|\frac{Y_r - \mathbb{1}[V_r > 0]}{f(V_r)}|^2$ is infinite, too. Therefore, for both cases, $\tilde{\psi} \notin L^2(P_\lambda)$.

If $\tilde{\psi}$ is the unique solution to $\langle A^*\psi, (h, \eta, g) \rangle = h$, then we have shown that $\partial_t \chi(\lambda_t) \notin R(A^*)$, in which $R(A^*)$ denotes $A^*$'s range and $A^*$ is the adjoint of the score operator $A$. $A^*$, $A$, $\chi(\lambda_t)$, and $\lambda_t$ are all defined in the proof of Corollary 4.3.1. Then by Theorem 25.32 of Van der Vaart (1998), we can conclude that there is no regular estimator in existence.

What left is to show that $\tilde{\psi}$ is indeed the unique solution to $\langle A^*\psi, (h, \eta, g) \rangle = h$. We assume there exists $\hat{\psi}(V, Y)$ which can also solve

$$
\langle A^*\psi, (h, \eta, g) \rangle = h \tag{C.0.15}
$$

Then, let $\pi(V, Y) := \hat{\psi}(V, Y) - \tilde{\psi}(V, Y)$. We aim to show $\pi(V, Y) \equiv 0$.

First, (C.0.15) implies that

$$
E\pi(V, Y)\eta(\varepsilon) = 0, \quad \forall \eta \in \phi P_\lambda,
$$

274
\( E\pi(V,Y)g(V) = 0, \quad \forall g \in \hat{\mathcal{P}}_{H_i}, \)

and

\[ E\dot{l}_\alpha(V,Y)\pi(V,Y) = 0. \]

Thus there exists some constants \( C_1 \) and \( C_2 \), such that

\[ E(\pi(V,Y)|\varepsilon) = C_1 + C_2\varepsilon, \quad E(\pi(V,Y)|V) = 0, \quad \text{and} \quad E\dot{l}_\alpha(V,Y)\pi(V,Y) = 0. \]

Note that

\[ E(\pi(V,Y)|\varepsilon) = \int^{\varepsilon}_{-\alpha} \pi(v,1) f_{\varepsilon}(v) dv + \int^{\varepsilon}_{-\infty} \pi(v,0) f_{\varepsilon}(v) dv, \]

in which \( f_{\varepsilon}(\cdot) \) is the density of \( V \). Taking derivatives on both sides w.r.t. \( \varepsilon \) and letting \( \varepsilon \) range over \( \mathbb{R} \), we have \( \pi(t,0) - \pi(t,1) = \frac{C_2}{f_{\varepsilon}(t)} \) for any \( t \in \mathbb{R} \). Then

\[ 0 = E\dot{l}_\alpha(V,Y)\pi(V,Y) = E\phi(V + \alpha)(\pi(V,1) - \pi(V,0)) = C_2 E\frac{\phi(V + \alpha)}{f_{\varepsilon}(V)} = C_2. \]

This implies \( \pi(t,1) = \pi(t,0) := \pi(t) \). At last,

\[ 0 = E(\pi(V,Y)|V) = \pi(V), \]

i.e. \( \pi(V,Y) \equiv 0 \). This concludes the uniqueness of \( \tilde{\psi} \) and thus the whole proof.

\[ C.0.5 \quad \text{Proof of Theorem 4.4.1} \]

We begin with \( \hat{\Phi} \) which is defined in (4.4.3).

\[
\hat{\Phi} = \frac{1}{n} \sum_{i=1}^{n} Z_i(Y_i - 1\{V_i > 0\}) f(U_i) I_{n,i} + \frac{1}{n} \sum_{i=1}^{n} Z_i(Y_i - 1\{V_i > 0\}) f^2(U_i) I_{n,i}(f(U_i) - \hat{f}(U_i)) + R_{n,1} + R_{n,2} + R_{n,3}
\]
in which \( \hat{U}_i = V_i - S_i \gamma \),

\[
R_{n,1} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{Z_i(Y_i - 1\{V_i > 0\})}{f(U_i)} \right) (\hat{I}_{n,i} - I_{n,i}),
\]

\[
R_{n,2} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{Z_i(Y_i - 1\{V_i > 0\})}{f(U_i)} \right) \left( \frac{f(U_i) - \hat{f}(\hat{U}_i)}{\hat{f}(\hat{U}_i)} \right) (\hat{I}_{n,i} - I_{n,i}),
\]

\[
R_{n,3} = \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{Z_i(Y_i - 1\{V_i > 0\})}{f(U_i)} \right] \left[ \frac{(f(U_i) - \hat{f}(\hat{U}_i))^2}{f(U_i) \hat{f}(\hat{U}_i)} \right] I_{n,i}.
\]

By Lemma C.0.12,

\[
\sum_{j=1}^{3} R_{n,j} = o_p\left( \frac{1}{\sqrt{n}} \right).
\]

Hence,

\[
\Phi = \frac{1}{n} \sum_{i=1}^{n} \frac{Z_i(Y_i - 1\{V_i > 0\})}{f(U_i)} I_{n,i} + \frac{1}{n} \sum_{i=1}^{n} \frac{Z_i(Y_i - 1\{V_i > 0\})}{f^2(U_i)} I_{n,i} (f(U_i) - \hat{f}(\hat{U}_i)) + o_p\left( \frac{1}{\sqrt{n}} \right).
\]

(C.0.16)

Next, we further decompose \( f(U_i) - \hat{f}(\hat{U}_i) \). Define \( \hat{f}(U_i) := \frac{1}{(n-1)h} \sum_{j \neq i} K(\frac{U_j - U_i}{h}) \) in which \( h \) is the tuning parameter defined in Assumption 39. Then we have

\[
f(U_i) - \hat{f}(\hat{U}_i)
\]

\[
= f(U_i) - \hat{f}(U_i) + \hat{f}(U_i) - \hat{f}(\hat{U}_i)
\]

\[
= f(U_i) - \frac{1}{(n-1)h} \sum_{j \neq i} K\left( \frac{U_j - U_i}{h} \right) + \frac{1}{(n-1)h} \sum_{j \neq i} (K\left( \frac{U_j - U_i}{h} \right) - K\left( \frac{\hat{U}_j - \hat{U}_i}{h} \right))
\]

\[
= f(U_i) - \frac{1}{(n-1)h} \sum_{j \neq i} K\left( \frac{U_j - U_i}{h} \right) + \frac{1}{(n-1)h^2} \sum_{j \neq i} K'\left( \frac{U_j - U_i}{h} \right) (U_j - U_i - (\hat{U}_j - \hat{U}_i))
\]

\[
+ \frac{1}{2(n-1)h^3} \sum_{j \neq i} K''\left( \frac{U_j - U_i}{h} \right) (U_j - U_i - (\hat{U}_j - \hat{U}_i))^2,
\]

where \( \hat{U}_j - \hat{U}_i \) is between \( \hat{U}_j - U_i \) and \( U_j - U_i \).
Since
\[
\max_{1 \leq i \leq n} |U_i - \hat{U}_i| = \max_{1 \leq i \leq n} |Z'(\hat{\gamma} - \gamma)| = O_p\left(\frac{1}{\sqrt{n}}\right)
\]
and \(h = n^{-H}\) for \(H < \frac{1}{6}\), we have
\[
\max_{i \leq n} \left| \frac{1}{2(n-1)h^3} \sum_{j \neq i} K''\left(\frac{\tilde{U}_j - \tilde{U}_i}{h}\right)(U_j - U_i - (\hat{U}_j - \hat{U}_i))^2 \right| \lesssim O_p\left(\frac{1}{nh^3}\right) = o_p\left(\frac{1}{\sqrt{n}}\right).
\]
This implies
\[
\left| \frac{1}{n} \sum_{i=1}^{n} \frac{Z_i(Y_i - 1\{V_i > 0\})}{f^2(U_i)} I_{n,i} \left(\frac{1}{2(n-1)h^3} \sum_{j \neq i} K''\left(\frac{\tilde{U}_j - \tilde{U}_i}{h}\right)(U_j - U_i - (\hat{U}_j - \hat{U}_i))^2 \right) \right| \lesssim \frac{1}{n} \sum_{i=1}^{n} \frac{|Y_i - 1\{V_i > 0\}|}{f^2(U_i)} I_{n,i} o_p\left(\frac{1}{\sqrt{n}}\right)
\]
\[
\lesssim o_p\left(\frac{1}{\sqrt{n}}\right)
\]
in which the last inequality is because \(\frac{1}{n} \sum_{i=1}^{n} \left| \frac{Y_i - 1\{V_i > 0\}}{f(U_i)} \right| = O_p(1)\) by Lemma C.0.14(2) or C.0.15(2). Then, we can simplify \((C.0.16)\) as
\[
\hat{\Phi} = \frac{1}{n} \sum_{i=1}^{n} \frac{Z_i(Y_i - 1\{V_i > 0\})}{f(U_i)} I_{n,i}
\]
\[
+ \frac{1}{n} \sum_{i=1}^{n} \frac{Z_i(Y_i - 1\{V_i > 0\})}{f^2(U_i)} I_{n,i} \left(\frac{1}{(n-1)h^2} \sum_{j \neq i} K'\left(\frac{U_j - U_i}{h}\right)(U_j - U_i - (\hat{U}_j - \hat{U}_i)) \right)
\]
\[
+ \frac{1}{n} \sum_{i=1}^{n} \frac{Z_i(Y_i - 1\{V_i > 0\})}{f^2(U_i)} I_{n,i} \left(\frac{1}{(n-1)h^2} \sum_{j \neq i} K'\left(\frac{U_j - U_i}{h}\right)(U_j - U_i - (\hat{U}_j - \hat{U}_i)) \right)
\]
\[
+ o_p\left(\frac{1}{\sqrt{n}}\right)
\]
\[
:= \delta_{n,1} + \delta_{n,2} + \delta_{n,3} + o_p\left(\frac{1}{\sqrt{n}}\right).
\]
277
For $\tilde{\delta}_{n,2}$, we will follow the same calculation of $\tilde{\delta}_{n,2}$ in the proof of Theorem 4.1 which we will not repeat. The key condition for applying the same argument is that $G(u) \in L^2(f^{1-\sigma}(u)du)$ where

$$G(u) = \frac{E(Z_i(Y_i - \mathbb{1}\{V_i > 0\})|U_i = u)}{f(u)} \mathbb{1}\{u \in S_n\}.$$ 

To see this, we note that

$$\left| \int G^2(u)f^{1-\sigma}(u)du \right| = \left| E\left(\frac{E(Z_i(Y_i - \mathbb{1}\{V_i > 0\})|U_i)}{f^{2+\sigma}(U_i)} \mathbb{1}\{U_i \in S_n\}\right) \right| < \infty,$$

in which the second last inequality is because $Z_i$ is bounded and the last inequality is by Lemma C.0.14(2) or C.0.15(2). Then we can obtain that

$$\tilde{\delta}_{n,2} = -\frac{1}{n} \sum_{i=1}^{n} \frac{E(Z_i(Y_i - \mathbb{1}\{V_i > 0\})|U_i)}{f(U_i)} I_{n,i} + E\left(\frac{E(Z_i(Y_i - \mathbb{1}\{V_i > 0\})|U_i)}{f(U_i)} \right) I_{n,i} + o_p\left(\frac{1}{\sqrt{n}}\right).$$

By Lemma C.0.14(1) or C.0.15(1), we have

$$E\left(\frac{E(Z_i(Y_i - \mathbb{1}\{V_i > 0\})|U_i)}{f(U_i)} \right) I_{n,i} = E\left(\frac{E(Z_i(Y_i - \mathbb{1}\{V_i > 0\})|U_i)}{f(U_i)} \right) + o\left(\frac{1}{\sqrt{n}}\right) = \Sigma x_\beta + o\left(\frac{1}{\sqrt{n}}\right).$$

So

$$\tilde{\delta}_{n,2} = -\frac{1}{n} \sum_{i=1}^{n} \frac{E(Z_i(Y_i - \mathbb{1}\{V_i > 0\})|U_i)}{f(U_i)} I_{n,i} + \Sigma x_\beta + o_p\left(\frac{1}{\sqrt{n}}\right). \quad (C.0.17)$$

Now let us turn to $\tilde{\delta}_{n,3}$ whose presence is due to the fact that $U_i$ is not directly observed. Let $W_i = (Y_i, S_i, U_i)$. Then we can write

$$\tilde{\delta}_{n,3} = \mathcal{U}_{n,3}(\hat{\gamma} - \gamma),$$

in which

$$\mathcal{U}_{n,3} = (C_n^2)^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} P_{n,3}(W_i, W_j).$$

278
\[ P_{n,3}(W_i, W_j) \]
\[
= \frac{1}{2} \left[ \frac{Z_i(Y_i - 1\{V_i > 0\})}{f^2(U_i)} I_{n,i} \frac{1}{h^2} K' \left( \frac{U_j - U_i}{h} \right) (S_j - S_i)' \right. \\
+ \left. \frac{Z_j(Y_j - 1\{V_j > 0\})}{f^2(U_j)} I_{n,j} \frac{1}{h^2} K' \left( \frac{U_i - U_j}{h} \right) (S_i - S_j)' \right].
\]

We next want to show that
\[ \mathcal{U}_{n,3} = -E \left( \frac{Z_i(Y_i - 1\{V_i > 0\}) f'(U_i)(ES_i - S_i)' \right) + o_p(1) \]  \hspace{1cm} (C.0.18)
\[ \text{and} \]
\[ E \left| \left( \frac{Z_i(Y_i - 1\{V_i > 0\}) f'(U_i)(ES_i - S_i)'}{f^2(U_i)} \right) \right| < \infty. \]  \hspace{1cm} (C.0.19)

Given (C.0.18), (C.0.19), and the fact that
\[ \hat{\gamma} - \gamma = \frac{1}{n} \sum_{i=1}^{n} \phi_i + o_p \left( \frac{1}{\sqrt{n}} \right), \]
in which \( \phi_i = \Sigma_{ss}^{-1} S_i U_i \), we have
\[ \tilde{\delta}_{n,3} = E \left( \frac{Z_i(Y_i - 1\{V_i > 0\}) f'(U_i)(S_i - ES_i)'}{f^2(U_i)} \right) \frac{1}{n} \sum_{i=1}^{n} \phi_i + o_p \left( \frac{1}{\sqrt{n}} \right). \]  \hspace{1cm} (C.0.20)

For (C.0.19), by Lemma C.0.14(2) or C.0.15(2), we have
\[ E \left| \left( \frac{Z_i(Y_i - 1\{V_i > 0\}) f'(U_i)(ES_i - S_i)'}{f^2(U_i)} \right) \right| \leq E \left| \frac{Y_i - 1\{V_i > 0\}}{f^2(U_i)} \right| < \infty. \]

To show (C.0.18), we first show that \( \text{Var}(\mathcal{U}_{n,3}) = o(1) \). This will imply that
\[ \mathcal{U}_{n,3} = E \mathcal{U}_{n,3} + o_p(1). \]  
To see that \( \text{Var}(\mathcal{U}_{n,3}) = o(1) \), we note that

\[
E|P_{n,3}(W_i, W_j)|^2 
\leq E \left[ Z_i(Y_i - \mathbb{1}\{V_i > 0\}) \right]^2 \frac{1}{f^4(U_i)} I_{n,3} \left( \frac{1}{h^4} E \left( |K'(\frac{U_j - U_i}{h})|^2 |S_j - S_i|^2 \right) W_i \right) 
\leq E \left[ Z_i(Y_i - \mathbb{1}\{V_i > 0\}) \right]^2 \frac{1}{f^4(U_i)} I_{n,3} \left( \frac{1}{h^3} \int |K'(\eta)|^2 f(U_i + h\eta) d\eta \right) 
\leq E \left[ Z_i(Y_i - \mathbb{1}\{V_i > 0\}) \right]^2 \frac{1}{f^{3+\sigma}(U_i)} I_{n,3} \left( \frac{1}{h^3} \right) 
\leq \frac{L_n}{h^3 n^{-\rho}} E \left[ \frac{|Y_i - \mathbb{1}\{V_i > 0\}|}{f^{2+\sigma}(U_i)} \right] 
\leq \frac{L_n}{h^3 n^{-\rho}}
\]

where the second last inequality is by Lemma C.0.13(1) and the last inequality is because of C.0.14(2) or C.0.15(2). In addition, we have

\[ 1 - \rho - 3H = 1 - 2\rho - 4H + \rho + H > 0. \]

This implies that

\[ \frac{L_n}{h^3 n^{-\rho}} = o(n) \]

and thus

\[ E|P_{n,3}(W_i, W_j)|^2 = o(n). \quad \text{(C.0.21)} \]

By Lemma A of Chapter 5 of Sackett et al. (2009), (C.0.21) implies the desired result that

\[ \text{Var}(\mathcal{U}_{n,3}) \leq \frac{2}{n} E|P_{n,3}(W_i, W_j)|^2 = o(1). \]
Next, we compute $EU_{n,3}$. Since $U \perp S$, we have $E(S_j | U_j) = E(S_j)$ and that

$$EU_{n,3} = EP_{n,3}(W_i, W_j)$$

$$= E \left[ \frac{Z_i(Y_i - 1\{V_i > 0\})}{f^2(U_i)} \right] I_{n,i} \left( \frac{ES_j - S_i}{h^2} \right) E \left( K'\left(\frac{U_j - U_i}{h}\right) | W_i \right).$$

First, we have that

$$\frac{1}{h^2} E \left( K'\left(\frac{U_j - U_i}{h}\right) | W_i \right) = \frac{1}{h} \int K'(\eta) f(U_i + h\eta) d\eta$$

$$= -f'(U_i) + R_n(U_i)$$

in which $|R_n(U_i)| \lesssim h^\nu$. Because $\sqrt{n}h^\nu \to \infty$ and

$$E \left| \frac{Z_i(Y_i - 1\{V_i > 0\})(ES_j - S_i)}{f^2(U_i)} \right| I_{n,i} < \infty,$$

$$EU_{n,3} = E \frac{Z_i(Y_i - 1\{V_i > 0\}) f'(U_i)(S_i - ES_j)' I_{n,i}}{f^2(U_i)} + o\left(\frac{1}{\sqrt{n}}\right).$$

In addition, we have

$$\frac{Z_i(Y_i - 1\{V_i > 0\}) f'(U_i)(S_i - ES_j)' I_{n,i}}{f^2(U_i)} \xrightarrow{p} \frac{Z_i(Y_i - 1\{V_i > 0\}) f'(U_i)(S_i - ES_j)' I_{n,i}}{f^2(U_i)}$$

and

$$\left| \frac{Z_i(Y_i - 1\{V_i > 0\}) f'(U_i)(S_i - ES_j)' I_{n,i}}{f^2(U_i)} \right| \leq \left| \frac{Y_i - 1\{V_i > 0\}}{f^2(U_i)} \right|.$$

So by Dominated Convergence Theorem,

$$E \frac{Z_i(Y_i - 1\{V_i > 0\}) f'(U_i)(S_i - ES_j)' I_{n,i}}{f^2(U_i)} = E \frac{Z_i(Y_i - 1\{V_i > 0\}) f'(U_i)(S_i - ES_j)'}{f^2(U_i)} + o(1)$$

and thus

$$U_{n,3} = EU_{n,3} + o_p(1) = E \frac{Z_i(Y_i - 1\{V_i > 0\}) f'(U_i)(S_i - ES_j)}{f^2(U_i)} + o_p(1).$$
This verifies (C.0.18). So (C.0.20) holds. Combining (C.0.16), (C.0.17) and (C.0.20), we have

\[
\hat{\Phi} - \Sigma_{xx}\beta = \frac{1}{n} \sum_{i=1}^{n} \frac{Z_i(Y_i - 1\{V_i > 0\})}{f(U_i)} I_{n,i} - \frac{1}{n} \sum_{i=1}^{n} \frac{E(Z_i(Y_i - 1\{V_i > 0\})|U_i)}{f(U_i)} I_{n,i} \\
+ E \left( \frac{Z_i(Y_i - 1\{V_i > 0\})f'(U_i)(S_i - ES_i)'}{f^2(U_i)} \right) \frac{1}{n} \sum_{i=1}^{n} \phi_i + o_p\left( \frac{1}{\sqrt{n}} \right).
\]

By Lemma C.0.14(1) or C.0.15(1) and the Markov inequality,

\[
\frac{1}{n} \sum_{i=1}^{n} \frac{Z_i(Y_i - 1\{V_i > 0\})}{f(U_i)} (1 - I_{n,i}) - \frac{1}{n} \sum_{i=1}^{n} \frac{E(Z_i(Y_i - 1\{V_i > 0\})|U_i)}{f(U_i)} (1 - I_{n,i}) = o_p\left( \frac{1}{\sqrt{n}} \right).
\]

Hence

\[
\hat{\Phi} - \Sigma_{xx}\beta = \frac{1}{n} \sum_{i=1}^{n} \Psi_i + o_p\left( \frac{1}{\sqrt{n}} \right)
\]

\[
\Psi_i = \frac{Z_i(Y_i - 1\{V_i > 0\})}{f(U_i)} - \frac{E(Z_i(Y_i - 1\{V_i > 0\})|U_i)}{f(U_i)}
\]

\[
+ E \left( \frac{Z_i(Y_i - 1\{V_i > 0\})f'(U_i)(S_i - ES_i)'}{f^2(U_i)} \right) \phi_i.
\]

(C.0.22)

Last, it is easy to see that \( E\Psi_i = 0 \) and

\[
E|\Psi_i|^{2+\sigma} \lesssim E \left| \frac{Z_i(Y_i - 1\{V_i > 0\})}{f(U_i)} \right|^{2+\sigma} + E|\phi_i|^{2+\sigma}
\]

\[
\lesssim E \left| \frac{Y_i - 1\{V_i > 0\}}{f^{2+\sigma}(U_i)} \right| + E|\phi_i|^{2+\sigma} < \infty.
\]

This implies that the Lindeberg condition holds. Then we have

\[
\sqrt{n}(\hat{\beta} - \beta) = (\Sigma'_{XX}W\Sigma_{xx})^{-1}\Sigma'_{XX}W\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\Psi_i - (Z_iX'_i - \Sigma_{xx})\beta) + o_p(1) \rightsquigarrow \mathcal{N}(0, \Sigma_\beta)
\]

in which \( \Sigma_0 = E(\Psi_i - (Z_iX'_i - \Sigma_{xx})\beta)(\Psi_i - (Z_iX'_i - \Sigma_{xx})\beta)' \) and

\[
\Sigma_\beta = (\Sigma'_{XX}W\Sigma_{xx})^{-1}\Sigma'_{XX}W\Sigma_0W\Sigma_{xx}^{-1}(\Sigma'_{XX}W\Sigma_{xx})^{-1}.
\]
C.0.6 Lemmas for Theorem 4.3.1

Note that \( \hat{I}_{n,i} := 1\{V_i \in \hat{S}_n\} \) in which \( \hat{S}_n := (\hat{l}_n, \hat{r}_n) \). The unusual feature of our trimming function is that the two endpoints of the interval are random. In order to deal with the randomness, we next propose two non-random intervals \( S^-_n \) and \( S^+_n \) such that \( S^-_n \subset \hat{S}_n \subset S^+_n \) w.p.a.1. Then we can bound the remainder terms using either of the two non-random intervals depending on the direction of the inequality.

We define

\[
S^+_n = (-M_{n,l} + l_n, M_{n,r} + r_n) \quad \text{and} \quad S^-_n = (M_{n,l} + l_n, -M_{n,r} + r_n)
\]

in which the two sequences \( M_{n,r} \) and \( M_{n,l} \) are chosen in Lemma C.0.6. Then by letting

\[
A_n = \{ |\hat{r}_n - r_n| \leq M_{n,r} \} \cap \{ |\hat{l}_n - l_n| \leq M_{n,l} \},
\]

Lemma C.0.6(1) shows that \( P(A_n) \to 1 \) and on \( A_n \),

\[
\hat{I}_{n,i} \leq 1\{V_i \in S^+_n\},
\]

\[
|\hat{I}_{n,i} - I_{n,i}| \leq 1\{V_i \in S^+_n\} - 1\{V_i \in S^-_n\} \leq \min\left(1 - 1\{V_i \in S^-_n\}, 1\{V_i \in S^+_n\}\right).
\]

(C.0.23)

We can derive bounds for various terms by replacing the random interval \( \hat{S}_n \) with two non-random intervals \( S^-_n \) and \( S^+_n \). This has been done in Lemma C.0.8 – C.0.11. These bounds are applied to derive the desired results in Lemma C.0.5.

Lemma C.0.5. Under the conditions of Theorem 4.3.1, we have

\[
R_{n,1} + R_{n,2} + R_{n,3} = o_p\left(\frac{1}{\sqrt{n}}\right).
\]
Proof. First, on \( A_n \), we have

\[
|R_{n,1}| \leq \frac{1}{n} \sum_{i=1}^{n} \left| \frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)} \right| (1 - \mathbb{1}\{S_n^-\}).
\]

Lemma C.0.8(1) or C.0.9(1) shows that

\[
E \left| \frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)} \right| (1 - \mathbb{1}\{S_n^-\}) = o\left( \frac{1}{\sqrt{n}} \right).
\]

Hence, by Markov inequality, we have

\[
R_{n,1} = o_p\left( \frac{1}{\sqrt{n}} \right).
\]

Similarly, on \( A_n \), we have

\[
|R_{n,2}| \leq \frac{1}{n} \sum_{i=1}^{n} \left| \frac{Y_i - \mathbb{1}\{V_i > 0\}}{f(V_i)} \right| (1 - \mathbb{1}\{V_i \in S_n^-\}) \left| \frac{f(V_i) - \hat{f}(V_i)}{\hat{f}(V_i)} \right| \mathbb{1}\{V_i \in S_n^+\}.
\]

Lemma C.0.11 shows that

\[
\max_{1 \leq i \leq n} |f(V_i) - \hat{f}(V_i)| \mathbb{1}\{V_i \in S_n^+\} \leq \sup_{v \in S_n^+} \left| \frac{1}{n-1} \left( n\hat{f}(v) - \frac{1}{h} K\left( \frac{v-V_i}{h} \right) \right) - f(v) \right|
\]

\[
\leq \frac{n}{n-1} \sup_{v \in S_n^+} |\hat{f}(v) - f(v)| + O\left( \frac{1}{nh} \right)
\]

\[
= O_p\left( \left( \frac{\log(n)}{nh} \right)^{\frac{1}{2}} \right). \tag{C.0.24}
\]

Lemma C.0.6(2) shows that, for \( \rho = \min(\rho_l(1 + \zeta_l), \rho_r(1 + \zeta_r)) \),

\[
|\hat{f}(V_i)| \mathbb{1}\{V_i \in S_n^+\} \geq (f(V_i) - \frac{n}{n-1} \sup_{v \in S_n} |\hat{f}(v) - f(v)| - O\left( \frac{1}{nh} \right)) \mathbb{1}\{V_i \in S_n^+\}
\]

\[
\geq c(f(V_i) + O_p\left( \left( \frac{\log(n)}{nh} \right)^{\frac{1}{2}} \right)) \mathbb{1}\{V_i \in S_n^+\}
\]

\[
\geq cn^{-\rho} L_n \mathbb{1}\{V_i \in S_n^+\}.
\]

284
Combining the above two inequalities with the fact that $1 - 2\rho - H > 0$, we obtain

$$\max_{1 \leq i \leq n} \left| \frac{f(V_i) - \hat{f}(V_i)}{f(V_i)} \right| 1\{V_i \in S_n^+\} = O_p\left( \frac{L_n}{\sqrt{n^{1-2\rho} h}} \right) = o_p(1).$$

Hence

$$|R_{n,2}| \leq \left| \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i - 1\{V_i > 0\}}{f(V_i)} \right| (1 - 1\{V_i \in S_n^-\})o_p(1) = |R_{n,1}|o_p(1) = o_p\left( \frac{1}{\sqrt{n}} \right).$$

For $R_{n,3}$, we have

$$\sqrt{n}|R_{n,3}| \leq \left| \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i - 1\{V_i > 0\}}{f(V_i)} \right| \left[ \sqrt{n} \left( \frac{f(V_i) - \hat{f}(V_i))^2}{f^2(V_i)(1 + o_p(1))} I_{n,i} \right) \right].$$

If Assumption 32(1) or (3) holds, then by Lemma C.0.8, we have

$$\frac{1}{n} \sum_{i=1}^{n} \left| \frac{Y_i - 1\{V_i > 0\}}{f^3(V_i)} \right| = O_p(1)$$

and

$$\max_{1 \leq i \leq n} \left| \sqrt{n}(f(V_i) - \hat{f}(V_i))^2 I_{n,i} \right| = O_p\left( \frac{\log(n)}{\sqrt{n} h} \right) = o_p(1).$$

This is sufficient to conclude that $R_{n,3} = o_p\left( \frac{1}{\sqrt{n}} \right)$.

On the other hand, if Assumption 32(2) holds, then by Lemma C.0.6(2) and the fact that $1 - 2\rho - 2H > 0$, we have

$$\frac{1}{n} \sum_{i=1}^{n} \left| \frac{Y_i - 1\{V_i > 0\}}{f(V_i)} \right|^2 = O_p(1)$$

and

$$\max_{1 \leq i \leq n} \left| \sqrt{n}(f(V_i) - \hat{f}(V_i))^2 I_{n,i} \right| = O_p(n^{\mu-2} h^{-1} L_n) = o_p(1).$$

This is also sufficient for $R_{n,3} = o_p\left( \frac{1}{\sqrt{n}} \right)$. \qed
Lemma C.0.6. For \( l_n, r_n, \hat{l}_n \) and \( \hat{r}_n \) defined in the Section 4.3, if Assumption 29 holds, then there exists two positive sequences \( M_{n,r} \) and \( M_{n,l} \) which satisfy the following conditions.

1. Let \( S_n^+ = (-M_{n,l} + l_n, M_{n,r} + r_n) \), \( S_n^- = (M_{n,l} + l_n, -M_{n,r} + r_n) \), and

\[
A_n = \{|\hat{r}_n - r_n| \leq M_{n,r} \} \cap \{|\hat{l}_n - l_n| \leq M_{n,l} \}.
\]

Then \( P(A_n) \to 1 \) and on \( A_n \), \( S_n^- \subset \hat{S}_n \subset S_n^+ \).

2. On \( S_n^+ \), \( f(v) \geq cn^{-\rho}L_n \) for \( \rho = \min(\rho_l(1 + \zeta_l), \rho_r(1 + \zeta_r)) \).

3. \( \exists B > 0 \) such that \( \frac{M_{n,r}}{r_n} \to 0 \), \( \frac{M_{n,l}}{l_n} \to 0 \), \( r_n + M_{n,r} \leq n^B \) and \( l_n + M_{n,l} \leq n^B \).

Lemma C.0.6(1) states that the feasible random interval \( \hat{S}_n \) nests and is nested by two deterministic intervals \( S_n^- \) and \( S_n^+ \) respectively w.p.a.1. Lemma C.0.6(2) shows that, on interval \( S_n^+ \), the decay rate of the density \( f(v) \) is controlled by tuning parameters \( \rho_r \) and \( \rho_l \). Lemma C.0.6(3) implies that the interval \( S_n^+ \) expands in polynomial of \( n \) which is needed when proving the uniform convergence rate of the density estimator with compact but expanding support.

Before showing the proof, we first introduce a lemma derived by Dekkers and De Haan (1989) on the asymptotic properties of extremal quantile estimators \( \hat{l}_n \) and \( \hat{r}_n \). Recall that \( F \) is the c.d.f. of special regressor \( V \). Let \( U = (\frac{1}{1-F})^{-}, \ V(t) = U(e^t) \).

Then we have \( V'(t) = U'(e^t)e^t \). By Assumption 31, \( U'(t) \in RV_{\zeta,-1} \). Denote \( E_{(1)}^{(n)} \leq E_{(2)}^{(n)} \leq \cdots \leq E_{(n)}^{(n)} \) the ascending order statistics of \( E_1, E_2, \cdots E_n \) where \( E_1, E_2, \cdots E_n \) are i.i.d. standard exponential random variables.

Lemma C.0.7. For \( m(n) \to \infty \) and \( m = \lfloor m(n) \rfloor \), we have

1. \( \{V_{(n-i+1)}^{(n)}, i = 1, 2, \cdots n\} \overset{d}{=} \{V(E_{(n-i+1)}^{(n)}), i = 1, 2, \cdots n\} \).
(2) If $\frac{m(n)}{n} \to 0$, $m(n) \to \infty$, let $\hat{r}_n = V_{(n-m+1)}^n$, then

$$
\sqrt{2m} \left( \frac{\hat{r}_n - r_n}{V_{(n-m+1)}^n - V_{(n-2m+1)}^n} \right) \rightsquigarrow \mathcal{N}(0, \sigma^2(\zeta_r))
$$

where $\sigma(\zeta_r)$ is a constant only depends on $\zeta_r$.

(3)

$$
\sqrt{2m} \left( \frac{V(E_{(n-m+1)}^n) - V(E_{(n-2m+1)}^n)}{2 \zeta_r V'(E_{(n-2m+1)}^n)} - \frac{1 - 2^{-\zeta_r}}{\zeta_r} \right) \rightsquigarrow \mathcal{N}(0, 1).
$$

(4)

$$
\frac{V'(E_{(n-2m+1)}^n)}{V'(\log(\frac{n}{2m})]} - 1 = \zeta_r \frac{V(E_{(n-2m+1)}^n) - V(\log(\frac{n}{2m}))}{V'(\log(\frac{n}{2m}))} + o_p\left( \frac{1}{\sqrt{2m}} \right) = O_p\left( \frac{1}{\sqrt{2m}} \right) = o_p(1).
$$

In our definition, for the right tail $m_r = n^{1-\rho_r}$ for some $0 < \rho_r < 1$. The convergence rate for $\hat{r}_n - r_n$ is

$$
\frac{\sqrt{2m_r}}{V_{(n-m_r+1)}^n - V_{(n-2m_r+1)}^n} \sim C \frac{\sqrt{2m_r}}{V'(E_{(n-2m_r+1)}^n)}
$$

$$
\sim C \frac{\sqrt{2m_r}}{V'(\log(\frac{n}{2m_r}))},
$$

where the first equivalence if by Lemma C.0.7(3) and the second equivalence is by Lemma C.0.7(4). Similarly, for the left tail, we let $m_l = n^{1-\rho_l}$. Then the convergence rate for $\hat{l}_n - l_n$ is

$$
\frac{\sqrt{2m_l}}{V'(\log(\frac{n}{2m_l}))}
$$

Now we are ready to prove Lemma C.0.6.
Proof. We only show the results for the right tail. The argument for the left tail is symmetric.

For (1), by Lemma C.0.7, we have \( \hat{r}_n - r_n = O_p(\sqrt{\frac{2m_r}{\log(\frac{1}{m_r})}}) \) where \( m_r = n^{1-\rho_r} \).

Let \( M_n \) be some deterministic sequence such that \( M_n \to \infty \). We define \( M_{n,r} = \frac{M_n \sqrt{\log(\frac{1}{m_r})}}{2m_r} \). Then we have \( |\hat{r}_n - r_n| = o_p(M_{n,r}) \).

For (2), because of the monotonicity of \( f \), for \( z \in S^+_n \), we have

\[
\frac{f(z)}{f(r_n)} \geq \frac{f(r_n + M_{n,r})}{f(r_n)} = \frac{f((1 - F)^{-\rho_r} (1 - F(r_n + M_{n,r})))}{f((1 - F)^{-\rho_r} (1 - F(r_n)))}.
\]

By Assumption 31, \( f((1 - F)^{-\rho_r}) \in RV_{\zeta_r+1}(0) \). In addition, since \( V \) has unbounded support, \( \zeta_r \geq 0 \) and \( \zeta_r + 1 \geq 1 \). If

\[
\frac{1 - F(r_n + M_{n,r})}{1 - F(r_n)} \to 1, \quad (C.0.25)
\]

then \( \frac{f(r_n + M_{n,r})}{f(r_n)} \to 1 \) and thus \( f(z) > C f(r_n) = C f((1 - F)^{-\rho_r} (n^{-\rho_r})) = C n^{-\rho_r (\zeta_r + 1)} L_n \).

Therefore we only need to verify Equation (C.0.25). The proof is divided into two cases.

Case (1): \( \zeta_r > 0 \), \( 1 - F \) is regularly varying. We only have to prove \( \frac{r_n + M_{n,r}}{r_n} \to 1 \) or equivalently, \( \frac{M_{n,r}}{r_n} \to 0 \). Note that by the choice of \( M_{n,r} \), we have

\[
\frac{M_{n,r}}{r_n} = M_n \frac{U'(n^{\rho_r}) n^{\rho_r}}{n^{1-\rho_r} (1 - F)^{-\rho_r} (n^{-\rho_r})} = \frac{M_n}{n^{1-\rho_r} L_n}.
\]

Since \( \rho_r < 1 \), the denominator diverges to infinity. Thus there exists a sequence \( M_n \) such that \( M_n \to \infty \) and \( \frac{M_{n,r}}{r_n} \to 0 \).
Case (2): $\zeta_r = 0$, then by Assumption 31, the right tail of $F$ is in the attraction domain of type 1 EV distribution. By Proposition 0.10 in Resnick (2007), it implies $\frac{1}{1-F}$ is $\Gamma$-varying with auxiliary function $f_0(t) = \frac{1-F(t)}{f(t)}$. Then $\frac{1-F(r_n+M_{n,r})}{1-F(r_n)} = \frac{1-F(r_n+M_{n,r})f_0(r_n)}{1-F(r_n)}$. If $\frac{M_{n,r}}{f_0(r_n)} \to 0$, then by the definition of $\Gamma$-varying function (Equation 0.47 in Resnick (2007)), (C.0.25) holds. Since $f(1-F)^{\leftarrow} \in RV_l(0)$,

$$f_0(r_n) = f_0((1-F)^{\leftarrow})(n^{-\rho_r}) \sim \frac{(1-F)(1-F)^{\leftarrow}(n^{-\rho_r})}{f(1-F)^{\leftarrow}(n^{-\rho_r})} \sim \frac{1}{L_n},$$

$f_0(r_n)$ is slowly varying. Therefore,

$$\frac{M_{n,r}}{f_0(r_n)} \sim \frac{M_n L_n}{n^{r_n^{-\rho_r}}}. \quad (C.0.26)$$

In addition, $\rho_r < 1$, we have $\frac{L_n}{n^{r_n^{-\rho_r}}} \to 0$. Thus there exists a sequence of $M_n$ such as $M_n = n^{r_n^{-\rho_r}} \to \infty$ such that $\frac{M_{n,r}}{f_0(r_n)} \to 0$. In addition, by (C.0.26), we have

$$\frac{M_{n,r}}{r_n} \sim \frac{M_n f_0(r_n) L_n}{r_n n^{r_n^{-\rho_r}}} \sim \frac{1}{r_n n^{r_n^{-\rho_r}} L_n} \to 0.$$

This is the desired result.

We can also use the similar argument to prove the statement for the left tail. Therefore, we conclude that on $S_n^+$, $f(V) \geq cn^{-\rho}L_n$, in which $\rho = \rho_r(\zeta_r+1)\vee \rho_l(\zeta_l+1)$.

For (3), notice that $\frac{M_{n,r}}{r_n} \to 0$ and if we have $E|V|^a < \infty$ for any $a > 0$, then $r_n = (1-F_V)^{\leftarrow}(n^{-\rho_l}) < n^{\frac{a}{\rho}}$. This implies $M_{n,r} + r_n \leq Cn^{\frac{a}{\rho}}$. \hfill $\square$

The next two lemmas verify the high level assumptions that ensure the $\sqrt{n}$-consistency of $\hat{\alpha}$: (1) the asymptotic bias vanishes faster than $\sqrt{n}$ and (2) the influence function’s $2+\sigma$ moment exists.

---

4 A non-decreasing function $U$ is $\Gamma$-varying if $U$ is defined on an interval $(x_l, x_0)$, $\lim_{x \to x_0} U(x) = \infty$ and there exists a positive function $f_0$ defined on $(x_l, x_0)$ such that for all $x$, $\lim_{t \to x_0} \frac{U(t+x f_0(t))}{U(t)} = \exp(x)$. 289
Lemma C.0.8. If Assumption 32(1) or (3) holds and the tuning parameters \( h \) and \( \rho_r \) are chosen as in Assumption 33(1), then the following statements hold.

(1) \( \sqrt{n}E|\frac{Y_i}{f_r(V_i)} - 1|^p \to 0 \) for any \( p > 0 \).

(2) \( E|\frac{Y_i}{f(V_i)} - 1|^p < \infty \) for any \( p > 0 \).

Proof. We only show the proof for the right tail. The proof for the left tail is symmetric. We note that, because the endpoint of the special regressor \( V \) is \( \pm \infty \), \( V \) can only have type 1 and type 2 tails, i.e. EV index is nonnegative. If \( V \) has type 1 tail, then \( (1 - F)^{\tau_r} \) is slowly varying and if it has type 3 tail, then \( (1 - F)^{\tau_r} \) is regularly varying.

For part (1), under Assumption 32(1), \( (1 - F)^{\tau_r}(z) \) is a regularly varying function and for any \( q_r > 0 \), \( (1 - F^\varepsilon)^{\tau_r}(z^{q_r}) \) is slowly varying as \( z \to 0 \). So we have \( \frac{(1-F)^{\tau_r}(z)}{(1-F^\varepsilon)^{\tau_r}(z^{q_r})} \to \infty \) as \( z \to 0 \). Under Assumption 32(3), for any \( q_r > 0 \), and as \( z \to 0 \),

\[
\frac{(1-F)^{\tau_r}(z)}{(1-F^\varepsilon)^{\tau_r}(z^{q_r})} = \frac{T_r^{\tau_r}(-z)}{D_r^{\tau_r}(-q_r \log(z))} = (-\log(z))^{\frac{1}{q_1} - \frac{1}{q_2}} L_n(-\log(z)) \to \infty.
\]

In addition, \( (1-F^\varepsilon)^{\tau_r}(z^{q_r}) \to \infty \) as \( z \to 0 \). Therefore, under either Assumption 32(1) or Assumption 32(3), for any \( q_r > 0 \) and any constant \( C \) independent of \( n \), when \( z \) is close enough to zero,

\[
\frac{C + (1-F)^{\tau_r}(z)}{(1-F^\varepsilon)^{\tau_r}(z^{q_r})} \geq 1. \tag{C.0.27}
\]

We define that

\[
B_{n,1} = \int_{r_n-m_{n,r}}^{\infty} \frac{1-F_\varepsilon(\alpha + v)}{f^{p-1}(v)} dv, \quad B_{n,2} = \int_{-\infty}^{-r_n \alpha + m_{n,l}} \frac{F_\varepsilon(\alpha + v)}{f^{p-1}(v)} dv.
\]
Then we have $E[Y_i - I[V_i > 0]] f(V_i) = B_{n,1} + B_{n,2}$. Note that similar to the proof of Lemma C.0.6, we have $\frac{1-F(r_n-M_{n,r})}{1-F(r_n)} \to 1$. Therefore, there exists a constant $C$ such that for $n$ large enough, we have

$$(1 - F(r_n - M_{n,r})) \leq C(1 - F(r_n)) = C((1 - F)((1 - F)^{\ast}(n^{-\rho_r}))) = Cn^{-\rho_r} L_n.$$  

(C.0.28)

(C.0.28) implies $\{z : (1 - F)^{\ast}(z) \geq r_n - M_{n,r}\} \subset \{z : z \leq Cn^{-\rho_r} L_n\}$. Let $v = (1 - F)^{\ast}(z)$. By the change of variables, we have, for an arbitrary $q_r > 0$,

$$\sqrt{n} B_{n,1} \leq \sqrt{n} \int_0^{Cn^{-\rho_r} L_n} \frac{(1 - F_z)(\alpha + (1 - F)^{\ast}(z))}{(f^p(1 - F)^{\ast}(z))} dz$$

$$\leq \sqrt{n} \int_0^{Cn^{-\rho_r} L_n} \frac{z^{q_r}}{(f^p(1 - F)^{\ast}(z))} dz$$

$$\leq \sqrt{n} \int_0^{Cn^{-\rho_r} L_n} z^{q_r - p(\zeta_r + 1)} L(z) dz$$

$$= O(n^{\frac{1}{2} - \rho_r (q_r - p(\zeta_r + 1) + 1)} L_n),$$

in which $L(z)$ is slowly varying function. The second inequality is by Equation (C.0.27) and the third inequality is by Assumption 31. Since $q_r$ is arbitrary, we can choose it to be large enough so that $\frac{1}{2} - \rho_r (q_r - p(\zeta_r + 1) + 1) < 0$. This means $\sqrt{n} B_{n,1} = o(1)$. Similarly, we can shown that $\sqrt{n} B_{n,2} = o(1)$. This concludes lemma C.0.8(1).

For part (2), we note that

$$E[Y_i - I[V_i > 0]] f(V_i)^p \leq C \int F_z(\alpha + v)(1 - I\{v > 0\}) + I\{v > 0\} (1 - F_z(\alpha + v))dv$$

$$= C(\int_0^\infty \frac{1 - F_z(\alpha + v)}{f^{p-1}(v)} dv + \int_0^0 \frac{F_z(\alpha + v)}{f^{p-1}(v)} dv).$$

We now only consider the integrability at the right tail. Let $z = (1 - F)^{\ast}(v)$.  

291
By the change of variables, we have
\[
\int_0^\infty \frac{1 - F_\varepsilon(\alpha + v)}{f^{p-1}(v)} dv = \int_0^c \frac{1 - F_\varepsilon(\alpha + v)}{f^{p-1}(v)} dv + \int_c^\infty \frac{1 - F_\varepsilon(\alpha + v)}{f^{p-1}(v)} dv
\leq C + \int_0^c \frac{(1 - F_\varepsilon)(\alpha + (1 - F)^{\varepsilon^{-1}}(z))}{(f(1 - F)^{\varepsilon^{-1}}(z))^p} dz
\leq C + \int_0^c z^{q_r - p}\zeta_r^{-1} L(z) dz.
\]

Since \( q_r \) is arbitrary, we can choose \( q_r \) large enough so that \( q_r - p(\zeta_r + 1) + 1 > 0 \). This implies the integral is finite at 0. By applying the similar argument, we can show the integral at the left tail is finite too and therefore Lemma C.0.8(2) holds. □

**Lemma C.0.9.** If Assumption 32(2) holds and the tuning parameter \( h \) and \( \rho_r \) are chosen as in Assumption 33(2), then the following statements hold.

1. \( \sqrt{n}E \left[ \frac{\sum_{i=1}^n \mathbb{1}_{V_i > 0} f(V_i)}{n} \right] |(1 - \mathbb{1}_{V_i \in S_n^-})| \to 0. \)

2. For the \( \sigma > 0 \) in Assumption 29, \( E \left[ \frac{\sum_{i=1}^n \mathbb{1}_{V_i > 0} f(V_i)}{n} \right]^{2+\sigma} < \infty. \)

**Proof.** For part (1), we first let \( q_r \) be some positive constant such that \( q_r < \frac{\zeta_r}{\lambda_r} \). Then by Assumption 31(2), for any arbitrary constant \( C \),
\[
\frac{C + (1 - F)^{\varepsilon^{-1}}(z)}{(1 - F_\varepsilon)^{\varepsilon^{-1}}(z^{q_r})} \sim z^{q_r - \zeta_r} L(z) \to \infty
\]
as \( z \to \infty \) in which \( L(z) \) is a slowly varying function.

In addition, by Assumption 33(2), we have \( \frac{\lambda_r}{2\zeta_r(1-\lambda_r)} < \rho_r \) or equivalently, \( \rho_r (\frac{\zeta_r}{\lambda_r} - \zeta_r) > \frac{1}{2} \). It implies that the \( q_r \) we previously choose can further satisfy that \( \rho_r (q_r - \zeta_r) > \frac{1}{2} \). Therefore, the same calculation in the proof of Lemma C.0.8(1) with the new \( q_r \) and \( p = 1 \) leads to part (1).

For part (2), we note that
\[
\frac{\zeta_r}{\lambda_r} - (2 + \sigma)(\zeta_r + 1) + 1 = (\zeta_r + 1)\left(\frac{\zeta_r(1 - \lambda_r)}{\lambda_r(1 + \zeta_r)} - (1 + \sigma)\right) > 0,
\]

292
in which the last inequality is by Assumption 33(2). Therefore, it implies that the \( q_r \) we previously choose can further satisfy \( q_r - (2 + \sigma)(\zeta_r + 1) + 1 > 0 \) hold. Then the same argument in the proof of Lemma C.0.8(2) with the new \( q_r \) and \( p = 2 + \sigma \) leads to part (2).

**Lemma C.0.10.** Let \( G(v) = \frac{P(v) - \mathbb{1}\{v > 0\}}{f(v)} \mathbb{1}\{v \in S_u\} \) where \( P(v) = P(Y = 1|V = v) \) and \( f \) is the density of the special regressor \( V \). If Assumption 29(3) holds, then for any fix \( u \in [-1, 1] \) and \( h \to 0 \), we have

1. \( \int G^2(v - hu)f(v)dv \leq C E|Y - \mathbb{1}\{V > 0\}|^{2+\sigma} \leq C \).
2. \( \int (G(v - hu) - G(v))^2f(v)dv \to 0 \) as \( h \to 0 \).

**Proof.** For part (1), we have

\[
\int G^2(v - hu)f(v)dv = \int G^2(v)f(v + hu)dv \\
\leq C \int \left( \frac{P(v) - \mathbb{1}\{v > 0\}}{f(v)} \right)^2 f(v)^{1-\sigma}dv \\
= CE \left( E \frac{Y - \mathbb{1}\{V > 0\}|}{f(V)^{1+\frac{\sigma}{2}}} |V|^2 \right) \\
\leq CE \left( \frac{(Y - \mathbb{1}\{V > 0\})^2}{f(V)^{2+\sigma}} \right) \\
= CE \left[ \frac{Y - \mathbb{1}\{V > 0\}}{f(V)} \right]^{2+\sigma} \leq C,
\]

in which the first equality is by the change of variables and the fact that \( v \) has full support. The first inequality is by Assumption 29(3). The second inequality is by Jansen’s inequality. The third equality is because \( |Y - \mathbb{1}\{V > 0\}| \) only take value 0 or 1. The last inequality is by Lemma C.0.8(2) or C.0.9(2).

For part (2), we will follow the proof of Lemma (0.12) in Folland (1995). Because \( G(v) \in L^2(f^{1-\sigma}(v)dv) \), for any \( \delta > 0 \), we can pick a continuous function \( g \) with
compact support so that

$$\int (G(v) - g(v))^2 f^{1-\sigma}(v) \, dv < \delta.$$  

This implies

$$\int (G(v) - g(v))^2 f(v) \, dv < c_1 \delta$$

where $c_1 = \sup_v f(v)$. Second, because $g$ is continuous with compact support, we have

$$\int (g(v - hu) - g(v))^2 f(v) \, dv \leq \delta,$$

as $h \to 0$. Last, we have

$$\int (G(v - hu) - g(v - hu))^2 f(v) \, dv = \int (G(v) - g(v))^2 f(v + hu) \, dv$$

$$\leq c_2 \int (G(v) - g(v))^2 f(v)^{1-\sigma} \, dv$$

$$\leq c_2 \delta.$$  

Combining all three inequalities with the triangular inequality, we have

$$\int (G^2(v - hu) - G^2(v)) \, f(v) \, dv \leq (c_1 + 1 + c_2) \delta.$$  

Since $\delta$ is arbitrary, this concludes the proof. \qed

**Lemma C.0.11.** Under Assumption 28, 30 and 31, $\sup_{S_n} |\hat{f}(v) - f(v)| = O_p((\log(n) / nh)^{\frac{1}{2}} + h^\nu)$. By our choice of $h$ such that $nh^{2\nu+1} \to 0$, the leading term is $O_p((\log(n) / nh)^{\frac{1}{2}})$.

**Proof.** The proof generally follows the partition argument used for the usual proof of uniform consistency of the kernel density estimator over a fixed compact set (see Li and Racine (2011) section 1.12). In their proof, $l_n$, the length of the cube (interval
in our case because $V$ is a scalar) is $\frac{\log(n)\frac{3}{2} h\frac{3}{2}}{n^2}$ which makes their Equations (1.61) and (1.62) hold. Given the whole length of the support is less than $Cn^B$ by Lemma C.0.6(3), the total number of interval $L_n$ in our case is $\frac{Cn^B}{l_n}$ which is still polynomial in $n$. So by choosing an even larger $\alpha$ in their Equation (1.59), their Equation (1.60) still holds. \hfill \Box

### C.0.7 Lemmas for Theorem 4.4.1

**Lemma C.0.12.** The assumptions in Theorem 4.4.1 hold. Then, for $R_{n,j}$, $j = 1, 2, 3$ defined in the proof of Theorem 4.4.1, we have

$$\sum_{j=1}^{3} R_{n,j} = o_p\left(\frac{1}{\sqrt{n}}\right).$$

The idea of the proof is similar to the proof of Lemma C.0.5. The key here is to find two deterministic sequences $M_{n,r}$ and $M_{n,l}$ such that for

$$A_n = \{|\tilde{I}_n - l_n| \leq M_{n,l} \cap |\tilde{r}_n - r_n| \leq M_{n,r}\},$$

we have $P(A_n) \to 1$. This has been done in Lemma C.0.13.

Next, we define $S^+_n = (-M_{n,r} + l_n, r_n + M_{n,r})$, $S^-_n = (M_{n,l} + l_n, r_n - M_{n,r})$, $I^+_{n,i} = 1\{U_i \in S^+_n\}$, and $I^-_{n,i} = 1\{U_i \in S^-_n\}$. On $A_n$, we have

$$I^-_{n,i} < I_{n,i} < I^+_{n,i}$$

for all $i = 1, \cdots, n$. This implies

$$I^-_{n,i} - 1 \leq I^-_{n,i} - I_{n,i} \leq I_{n,i} - I_{n,i} \leq I^+_{n,i} - I_{n,i} \leq \min(1 - I^-_{n,i}, I^+_{n,i}). \tag{C.0.29}$$

We can derive bounds for various terms by replacing $\tilde{I}_{n,i}$ by the non-random upper and lower bounds. This has been done by Lemma C.0.14 - C.0.16.

Given Lemma C.0.13 - C.0.16, we next prove Lemma C.0.12.
Proof. By Lemma C.0.13(1), on $A_n$, we have
\[ |\tilde{I}_{n,i} - I_{n,i}| \leq 1 - 1\{U_i \in S_{n}^{-}\}, \]
and $P(A_n) \to 1$.

Then by Lemma C.0.14(1) or C.0.15(1), on $A_n$,
\[ E|R_{n,1}| \leq E\left| \frac{Y_i - 1\{V_i > 0\}}{f(U_i)} \right| (1 - 1\{U_i \in S_{n}^{-}\}) \left| \frac{f(U_i) - \hat{f}(\hat{U}_i)}{f(U_i)} \right| 1\{U_i \in S_{n}^{+}\} = o\left( \frac{1}{\sqrt{n}} \right). \]
This implies $R_{n,1} = o_p\left( \frac{1}{\sqrt{n}} \right)$. For $R_{n,2}$, on $A_n$, we have
\[ |R_{n,2}| \leq \frac{1}{n} \sum_{i=1}^{n} \left| \frac{Y_i - 1\{V_i > 0\}}{f(U_i)} \right| (1 - 1\{U_i \in S_{n}^{-}\}) \left| \frac{f(U_i) - \hat{f}(\hat{U}_i)}{f(U_i)} \right| 1\{U_i \in S_{n}^{+}\}. \]
By Lemma C.0.16, we have
\[ \left| \frac{f(U_i) - \hat{f}(\hat{U}_i)}{\hat{f}(\hat{U}_i)} \right| 1\{U_i \in S_{n}^{+}\} \leq O_p\left( \sqrt{\frac{L_n}{n^{1-2p}\Delta^4}} \right) 1\{U_i \in S_{n}^{+}\} \leq o_p(1), \]
in which the last inequality holds for all $i = 1, \cdots, n$. Furthermore, we have already show in Lemma C.0.14(1) or C.0.15(1) that
\[ \frac{1}{n} \sum_{i=1}^{n} \left| \frac{Y_i - 1\{V_i > 0\}}{f(U_i)} \right| (1 - 1\{U_i \in S_{n}^{-}\}) = o_p\left( \frac{1}{\sqrt{n}} \right). \]
This implies $R_{n,2} = o_p\left( \frac{1}{\sqrt{n}} \right)$. For $R_{n,3}$, on $A_n$, we have
\[ |R_{n,3}| \leq \frac{1}{n} \sum_{i=1}^{n} \left| \frac{Y_i - 1\{V_i > 0\}}{f(U_i)} \right| \left| \frac{\sqrt{n}(f(U_i) - \hat{f}(\hat{U}_i))^2}{f^2(U_i)(1 + o_p(1))} \right| 1\{U_i \in S_{n}\}. \]
If Assumption 38(1) or (3) hold, then by Lemma C.0.14(1),
\[ \frac{1}{n} \sum_{i=1}^{n} \left| \frac{Y_i - 1\{V_i > 0\}}{f(U_i)} \right| = O_p(1) \]
and by Lemma C.0.16

$$\sqrt{n}(f(U_i) - \hat{f}(\hat{U}_i))^2 = O_p\left(\frac{1}{\sqrt{n}h^4}\right) = o_p(1).$$

This concludes that \( R_{n,3} = o_p\left(\frac{1}{\sqrt{n}}\right). \)

On the other hand, if Assumption 38(2) hold, then by Lemma C.0.15(1),

$$\frac{1}{n} \sum_{i=1}^{n} \left| \frac{Y_i - 1\{V_i > 0\}}{f^2(U_i)} \right| = O_p(1)$$

and by Lemma C.0.16

$$\frac{\sqrt{n}(f(U_i) - \hat{f}(\hat{U}_i))^2}{f(U_i)} I_{n,i}^+ = O_p\left(\frac{L_n n^p}{\sqrt{n}h^4}\right) = o_p(1),$$

in which the last equality is because by Assumption 39(2),

$$1 - 8H - 2\rho > 0.$$  

Again in this case, \( R_{n,3} = o_p\left(\frac{1}{\sqrt{n}}\right). \)

\[\square\]

**Lemma C.0.13.** There exists positive sequences \( M_{n,l} \) and \( M_{n,l} \) such that for \( A_n = \left\{ \mid \tilde{l}_n - l_n \mid \leq M_{n,l} \cap \mid \tilde{r}_n - r_n \mid \leq M_{n,r} \right\} \), \( S_n^+ = (-M_{n,l} + l_n, r_n + M_{n,r}) \), and \( S_n^- = (M_{n,l} + l_n, r_n - M_{n,r}) \), we have

1. \( P(A_n) \to 1 \), and on \( A_n \), \( \{U_i \in S_n^-\} \subset \{\hat{U}_i \in \hat{S}_n\} \subset \{U_i \in S_n^+\} \) for \( i = 1, \ldots, n. \)
2. \( \forall u \in S_n^+, \ f(u) \geq cn^{-\rho}L_n \) where \( \rho = \min(\rho_r(1 + \zeta_r), \rho_l(1 + \zeta_l)). \)
3. \( \exists B > 0 \) such that \( \frac{M_{n,r}}{r_n} \to 0, \frac{M_{n,l}}{l_n} \to 0, \mid r_n \mid + M_{n,r} \leq n^B \), and \( \mid l_n \mid + M_{n,l} \leq n^B. \)

**Proof.** The only difference between Lemma C.0.13 and Lemma C.0.6 is that now \( U_i \) is unobservable and we will replace it by the residual \( \hat{U}_i \). Then the feasible trimming points \( \tilde{l}_n \) and \( \tilde{r}_n \) are computed as order statistics of \( \hat{U}_i \). By Assumption 36 and the fact that \( \hat{\gamma} \) is \( \sqrt{n}\)-consistent, we have

$$\max_{1 \leq i \leq n} |\hat{U}_i - U_i| \leq \max_{1 \leq i \leq n} |S_i| |\hat{\gamma} - \gamma| = O_p\left(\frac{1}{\sqrt{n}}\right).$$
Since the convergence rate for intermediate order statistics is slower than $\sqrt{n}$, we expect that replacing $U_i$ by its estimator $\hat{U}_i$ will not change the asymptotic properties of the estimator. In particular, for left tail, let $\tau_{n,l} = n^{-\rho_l}$, $m_l = \lceil n\tau_{n,l} \rceil$, $\alpha_{n,l} = \sqrt{n} \tau_{n,l} f(F^{(r)}(\tau_{n,l}))$. We can show that

$$\hat{U}^{(n)}_{(m_l)} - l_n = O_p\left(\frac{1}{\alpha_{n,l}}\right).$$

(C.0.30)

Similarly, for the right tail, we can show that

$$\hat{U}^{(n)}_{(n-m_r+1)} - r_n = O_p\left(\frac{1}{\alpha_{n,r}}\right),$$

(C.0.31)

in which $\tau_{n,r} = n^{-\rho_r}$, $m_r = \lceil n\tau_{n,r} \rceil$, $\alpha_{n,r} = \sqrt{n} \tau_{n,r} f((1 - F)^{*(r)}(\tau_{n,r}))$. Given (C.0.30) and (C.0.31), for any sequence $M_n \to \infty$, we have

$$\frac{M_n \sqrt{n}}{\alpha_{n,r}} \sim \frac{M_n}{\tau_{n,r}^{\frac{1}{2} + \epsilon_r} L_n} \to \infty, \quad \frac{M_n \sqrt{n}}{\alpha_{n,l}} \sim \frac{M_n}{\tau_{n,l}^{\frac{1}{2} + \epsilon_l} L_n} \to \infty.$$

Then w.p.a.1,

$$\{\hat{U}_i \leq \tilde{r}_n\} \subseteq \{\hat{U}_i \leq r_n + |r_n - \tilde{r}_n|\} \subseteq \{U_i \leq r_n + \frac{M_n}{\alpha_{n,r}} + \max_{i \leq n} |U_i - \tilde{U}_i|\} \subseteq \{U_i \leq r_n + \frac{M_n}{\alpha_{n,r}} + C\sqrt{n}\} \subseteq \{U_i \leq r_n + 2M_n\}.$$

Similarly,

$$\{\hat{U}_i \leq \tilde{r}_n\} \supseteq \{U_i \leq r_n - 2M_n\}.$$

This implies we can choose $M_{n,r} = \frac{2M_n}{\alpha_{n,r}}$ for any $M_n \to \infty$ for the right tail. For the left tail, we can choose $M_{n,l} = \frac{2M_n}{\alpha_{n,l}}$. Then, by the same argument, for any $i = 1, \ldots, n,$
we have, w.p.a.1,
\[ \{ U_i \leq l_n + \frac{2M_n}{\alpha_{n,l}} \} \subset \{ \hat{U}_i \leq \hat{l}_n \} \subset \{ U_i \leq l_n - \frac{2M_n}{\alpha_{n,l}} \}. \]

Combining all the results together, we have shown part (1) of the lemma with \( S_n^+ = (-M_{n,l} + l_n, r_n + M_{n,r}) \) and \( S_n^- = (M_{n,l} + l_n, r_n - M_{n,r}) \). Part (2) and (3) follow the same argument of Lemma C.0.6 which we will not repeat. Now, the only pieces left unjustified are (C.0.30) and (C.0.31) which we will turn to next.

We focus on the left tail and show (C.0.30). (C.0.31) can be derived in the same manner. Let \( \rho_{\tau}(u) \) be the check function defined as \( \rho_{\tau}(u) = u(\tau - 1\{u \leq 0\}) \). Then \( \hat{U}_{(m_l)} \) can be viewed as one point in the following argmin set:

\[ \arg\min_q \sum_{i=1}^n \rho_{\tau_{n,l}}(\hat{U}_i - q). \]

Define \( z := \alpha_{n,l}(q - l_n) \), \( \hat{Z}_n := \alpha_{n,l}(\hat{U}_{(m_l)} - l_n) \), and

\[ Q(\tau_{n,l}, z) := \frac{\alpha_{n,l}}{\sqrt{n\tau_{n,l}}} \sum_{i=1}^n [\rho_{\tau_{n,l}}(U_i - l_n - \frac{z}{\alpha_{n,l}} + \hat{U}_i - U_i) - \rho_{\tau_{n,l}}(U_i - l_n)]. \]

Then \( \hat{Z}_n \) minimize \( Q(\tau_{n,l}, z) \).

In the following, we show that, \( Q(\tau_{n,l}, z) \), the rescaled version of the objective function, weakly converges to a limiting process

\[ -zW + \frac{z^2}{2} \]

where \( W \sim \mathcal{N}(0, 1) \). Then we can apply the Convexity lemma and the same argument in the proof of Theorem 1 of Pollard (1991b) to derive the desired result that

\[ \hat{Z}_n \rightsquigarrow W. \]
By equation (9.44) of Chernozhukov (2005c),
\[
Q(\tau_{n,l}, z) = \frac{-z}{\sqrt{n\tau_{n,l}}} \sum_{i=1}^{n} (\tau_{n,l} - 1\{U_i \leq l_n\}) + \frac{\alpha_{n,l}}{\sqrt{n\tau_{n,l}}} \sum_{i=1}^{n} (U_i - \hat{U}_i)(\tau_{n,l} - 1\{U_i \leq l_n\}) \\
+ \frac{\alpha_{n,l}}{\sqrt{n\tau_{n,l}}} \sum_{i=1}^{n} \int_{0}^{s_n,U_i} \left[1\{U_i - l_n \leq s\} - 1\{U_i - l_n \leq 0\}\right] ds.
\]
(C.0.32)

We next want to show that
\[
\frac{1}{\sqrt{n\tau_{n,l}}} \sum_{i=1}^{n} (\tau_{n,l} - 1\{U_i \leq l_n\}) \rightsquigarrow W, \quad (C.0.33)
\]
\[
\frac{\alpha_{n,l}}{\sqrt{n\tau_{n,l}}} \sum_{i=1}^{n} (U_i - \hat{U}_i)(\tau_{n,l} - 1\{U_i \leq l_n\}) = o_p(1), \quad (C.0.34)
\]
and
\[
\frac{\alpha_{n,l}}{\sqrt{n\tau_{n,l}}} \sum_{i=1}^{n} \int_{0}^{s_n,U_i} \left[1\{U_i - l_n \leq s\} - 1\{U_i - l_n \leq 0\}\right] ds \overset{p}{\longrightarrow} \frac{z^2}{2}. \quad (C.0.35)
\]

(C.0.33) holds because by the triangular array CLT such as Theorem 3.4.5 in Durrett (2010),
\[
\frac{1}{\sqrt{n\tau_{n,l}}} \sum_{i=1}^{n} (\tau_{n,l} - 1\{U_i \leq l_n\}) \rightsquigarrow \mathcal{N}(0,1).
\]
Here we note that the Lyapunov condition for the CLT holds because \(n\tau_n \to \infty\).

For (C.0.34), we have
\[
\frac{\alpha_{n,l}}{\sqrt{n\tau_{n,l}}} \sum_{i=1}^{n} (U_i - \hat{U}_i)(\tau_{n,l} - 1\{U_i \leq l_n\}) = \left[ \frac{1}{\sqrt{n\tau_{n,l}}} \sum_{i=1}^{n} (\tau_{n,l} - 1\{U_i \leq l_n\}) Z_i \right] [\alpha_{n,l}(\hat{\gamma} - \gamma)] \\
= o_p \left( \frac{\alpha_{n,l}}{\sqrt{n}} \right) = o_p \left( L(\tau_{n,l}) \tau_{n,l}^{x_1 + 0.5} \right) = o_p(1),
\]

300
in which the second last equality is by Assumption 31 and $L(\tau)$ is a slowly varying function at 0 such that $\frac{L(k\tau)}{L(\tau)} \to 1$ for any $k > 0$ as $\tau \to 0$.

Before proving (C.0.35), we first define $\hat{d} = \sqrt{n}(\hat{\gamma} - \gamma)$. Then $U_i - \hat{U}_i = Z_i \frac{\hat{d}}{\sqrt{n}}$.

Since $\hat{\gamma}$ is $\sqrt{n}$-consistent, $\hat{d} = O_p(1)$. Next we first consider

$$\Lambda_n(z, d) = \frac{\alpha_{n,l}}{\sqrt{\tau_{n,l}}} \sum_{i=1}^{n} \int_{0}^{\frac{z+\alpha_{n,l}Z_i d}{\sqrt{n}}} \left[ 1\{U_i - l_n \leq s\} - 1\{U_i - l_n \leq 0\} \right] ds$$

and show that $\Lambda_n(z, d) \overset{p}{\to} \frac{\hat{d}^2}{2}$ uniformly in $|z| \leq B, |d| \leq B$ for any $B > 0$. To see this, we note that

$$E\Lambda_n(z, d) = \frac{n}{\sqrt{\tau_{n,l}}} E \int_{0}^{\frac{z+\alpha_{n,l}Z_i d}{\sqrt{n}}} \left[ 1\{U_i - l_n \leq \frac{s}{\alpha_{n,l}}\} - 1\{U_i - l_n \leq 0\} \right] ds$$

$$= \frac{n}{\sqrt{\tau_{n,l}}} E \int_{0}^{\frac{z+\alpha_{n,l}Z_i d}{\sqrt{n}}} \left[ F(l_n + \frac{s}{\alpha_{n,l}}) - F(l_n) \right] ds$$

$$= \frac{n}{\sqrt{\tau_{n,l}}} E \int_{0}^{\frac{z+\alpha_{n,l}Z_i d}{\sqrt{n}}} f(l_n + \frac{\tilde{s}}{\alpha_{n,l}}) \frac{s}{\alpha_{n,l}} ds$$

in which $\tilde{s}$ is between 0 and $z + \frac{\alpha_{n,l}}{\sqrt{n}}Z_i d$. Since $|z| < B$, $|d| < B$, $|Z_i| < B$ and $\frac{\alpha_{n,l}}{\sqrt{n}} \to 0$, $\tilde{s}$ is bounded. Then by Equation (9.57) of Chernozhukov (2005c), we have

$$f(l_n + \frac{\tilde{s}}{\alpha_{n,l}}) \sim f(l_n).$$

Hence we have, uniformly over $z, d$

$$E\Lambda_n(z, d) \sim \sqrt{\frac{\tau_{n,l}}{n}} \frac{f(l_n)}{\alpha_{n,l}} \frac{1}{2} E(z + \frac{\alpha_{n,l}Z_i d}{\sqrt{n}})^2 \to \frac{s^2}{2}.$$
Next, we show
\[
\sup_{|d| < B, |z| < B} |\Lambda_n(z, d) - E\Lambda_n(z, d)| \to 0.
\]

Let us consider the class of functions
\[
\mathcal{F} = \left\{ \frac{1}{\sqrt{n}} \int_{0}^{z + Z_i \frac{d\alpha_{n,l}}{\sqrt{n}}} \left[ 1\{U_i - l_n \leq \frac{s}{\alpha_{n,l}}\} - 1\{U_i - l_n \leq 0\} \right] ds : |d| < B, |z| < B \right\}
\]
with envelope \( F_e = \frac{C}{\sqrt{n}} \). It is easy to see that \( \mathcal{F} \) satisfies the uniform entropy condition, that is,
\[
\sup_Q N(\varepsilon \|F_e\|_{Q, 2}, \mathcal{F}, \|\cdot\|_{Q, 2}) \leq \left( \frac{a}{\varepsilon} \right)^v, \quad \forall \varepsilon \in (0, 1].
\]

In addition, since \( Z_i \) is bounded, for \( \sigma_n^2 := \sup_{f \in \mathcal{F}} Ef^2 \), we have
\[
\sigma_n^2 \lesssim \tau_{n,l}^{-1} E \left[ 1\{U_i \leq l_n + \frac{B + |Z_i| B_{\alpha_{n,l}}}{\alpha_{n,l}}\} - 1\{U_i \leq l_n - \frac{B + |Z_i| B_{\alpha_{n,l}}}{\alpha_{n,l}}\} \right]
\]
\[
\lesssim \tau_{n,l}^{-1} f(l_n) \frac{1}{\alpha_{n,l}} = \frac{1}{\sqrt{n} \tau_{n,l}}
\]
in which the second inequality is by Equation (9.57) of Chernozhukov (2005c). Then by Corollary 5.1 of Chernozhukov et al. (2014), we have
\[
E \sup_{|d| < B, |z| < B} |\Lambda_n(z, d) - E\Lambda_n(z, d)|
\]
\[
= E\|\sqrt{n}(P_n - \mathcal{P})\|_F
\]
\[
\lesssim \sqrt{\frac{\sigma_n^2 \log(\frac{\|F_e\|_{P, 2}}{\sigma_n})}{\sqrt{n} \tau_{n,l}}} \log(\frac{\|F_e\|_{P, 2}}{\sigma_n})
\]
\[
\lesssim \sqrt{\frac{\log(n)}{\sqrt{n} \tau_{n,l}}} \to 0.
\]
This implies \( \sup\{d:|d|<B,|z|<B\} |\Lambda_n(z,d) - E\Lambda_n(z,d)| \xrightarrow{p} 0 \). Thus
\[
\Lambda_n(z,d) \xrightarrow{p} \frac{z^2}{2}
\]
uniformly in \(|z| < B, |d| < B\). Then uniformly over \(|z| < B\),
\[
\Lambda_n(z,d) \xrightarrow{p} \frac{z^2}{2}
\]
and thus (C.0.35) holds. Combining (C.0.33) – (C.0.35), we have
\[
Q(\tau_{n,l}, z) \xrightarrow{w} -zW + \frac{z^2}{2}.
\]
Since the RHS is uniquely minimized at \( z = W \), by the same argument of the proof of Theorem 1 in Pollard (1991b), we have
\[
\alpha_{n,l}(\hat{\gamma}_{(m)}^{(n)} - l_n) \xrightarrow{w} W \quad \text{and thus} \quad (\hat{\gamma}_{(m)}^{(n)} - l_n) = O_p\left(\frac{1}{\alpha_{n,l}}\right).
\]
This concludes the proof.

\begin{proof}
\end{proof}

\begin{Lemma}
If Assumption 38(1) or (3) holds and the tuning parameter \( h \) and \( (\rho_r, \rho_l) \) are chosen as in Assumption 39(1), then the following statements hold.
\begin{enumerate}
\item \( \sqrt{n}E \left| \frac{Y_i - 1\{V_i > 0\}}{f_p(U_i)} \right| (1 - 1\{U_i \in S_n^{-}\}) \).
\item \( E \left| \frac{Y_i - 1\{V_i > 0\}}{f_p(U_i)} \right| < \infty \) for any \( p > 0 \).
\end{enumerate}
\end{Lemma}

\begin{Proof}
For part (1), we have \( 1 - 1\{u \in S_n^{-}\} = 1\{u > r_n - M_{n,r}\} + 1\{u < l_n + M_{n,l}\} \).
Since \( S_i \) has bounded support, \( U_i > r_n - M_{n,r}, \) and \( \frac{M_{n,r}}{r_n} \to 0 \), we have \( V_i = U_i + S_i \gamma >
\end{Proof}
0. Similarly, \( U_i < l_n + M_{n,l} \) implies \( V_i = U_i + S_i^2 \gamma < 0 \). So we have

\[
\sqrt{n}E \left| \frac{Y_i - 1\{V_i > 0\}}{f^p(U_i)} \right| 1\{U_i > r_n - M_{n,r}\}
\]

\[
\leq \sqrt{n}E \left( \frac{1 - 1\{\varepsilon_i \leq X_i^p \beta + Z_i^p \gamma + U_i\}}{f^p(U_i)} \right) 1\{U_i > r_n - M_{n,r}\}
\]

\[
\leq \sqrt{n}E \left( \frac{1 - 1\{\varepsilon_i \leq U_i - C\}}{f^p(U_i)} \right) 1\{U_i > r_n - M_{n,r}\}
\]

\[
\leq \sqrt{n} \int_{r_n - M_{n,r}}^{\infty} \frac{(1 - F_u)(u - C)}{f^{p-1}(u)} du.
\]

Following the same argument, we have

\[
\sqrt{n}E \left| \frac{Y_i - 1\{V_i > 0\}}{f^p(U_i)} \right| 1\{U_i < l_n + M_{n,l}\} \leq \sqrt{n} \int_{-\infty}^{l_n + M_{n,l}} \frac{F_u(u + C)}{f^{p-1}(u)} du.
\]

The two terms on the RHS of the above two inequalities can be shown as \( o(1) \) in the same manner as \( B_{n,1} \) and \( B_{n,2} \) in the proof of Lemma C.0.8.

For part (2), we can choose \( C_1 \) such that when \( U_i > C_1, V_i > 0 \) and when \( U_i < -C_1, V_i < 0 \). On \( |U_i| \leq C_1 \), the integrand \( \frac{Y_i - 1\{V_i > 0\}}{f^p(U_i)} \) is bounded. So we only have to check the integrability at \( \pm \infty \). We focus on the right tail.

\[
E \left| \frac{Y_i - 1\{V_i > 0\}}{f^p(U_i)} \right| 1\{U_i > C_1\}
\]

\[
\leq \int_{C_1}^{\infty} \frac{(1 - F_u)(u - C)}{f^{p-1}(u)} du
\]

\[
\leq \int_{0}^{\infty} \frac{(1 - F_u)((1 - F)^{\gamma}(z) - C)}{f^p((1 - F)^{\gamma}(z))} dz < \infty.
\]

The last inequality holds for the same reason as part (2) of Lemma C.0.8.

\[ \square \]

**Lemma C.0.15.** If Assumption 38(2) holds and the tuning parameters \( h \) and \( (\rho_r, \rho_l) \) are chosen as in Assumption 39(2), then the following statements hold.
(1) $\sqrt{n}E\left|\frac{Y_i - 1\{Y_i > 0\}}{f(U_i)}\right| (1 - 1\{U_i \in S_n^-\})$.

(2) $E\left|\frac{Y_i - 1\{Y_i > 0\}}{f^{2+\sigma}(U_i)}\right| < \infty$.

Proof. For part (1), by repeating the proof of Lemma C.0.14 with $p = 1$, for the right tail, we have

$$\sqrt{n}E\left|\frac{Y_i - 1\{Y_i > 0\}}{f(U_i)}\right| 1\{U_i > r_n - M_{n,r}\} \leq \sqrt{n} \int_{r_n-M_{n,r}}^{\infty} (1 - F_z)(u - C) du.$$  

In order for the RHS to vanish, as in the proof of Lemma C.0.9, we need $\rho_r(\frac{\xi}{\lambda_r} - \zeta_r) > \frac{1}{2}$ which holds by Assumption 39(2).

For part (2), by repeating the proof of Lemma C.0.14 with $p = \sigma + 2$, for the right tail, we have

$$E\left|\frac{Y_i - 1\{Y_i > 0\}}{f^{2+\sigma}(U_i)}\right| 1\{U_i > C_1\} \leq \int_0^c \frac{(1 - F_z)((1 - F)^\sigma(z) - C)}{f^\sigma((1 - F)^\sigma(z))} dz. \quad (C.0.36)$$

Then, following the proof of Lemma C.0.9, the RHS of (C.0.36) is finite because $\frac{\xi}{\lambda_r} - (2 + \sigma)(\zeta_r + 1) + 1 > 0$ by Assumption 38(2). \hfill \Box

Lemma C.0.16.

$$\max_{1 \leq i \leq n} |f(U_i) - \hat{f}(\hat{U}_i)| 1\{U_i \in S_n^+\} = O_p\left(\frac{1}{\sqrt{nh^2}}\right)$$

and

$$|\hat{f}(\hat{U}_i)| 1\{U_i \in S_n^+\} \geq n^{-p} L_n |U_i \in S_n^+\}$$

where $p = \rho_r(1 + \zeta_r) \wedge \rho_l(1 + \zeta_l)$.

Proof. For the first result, we have $\max_{1 \leq i \leq n} |U_i - \hat{U}_i| \leq \max_{1 \leq i \leq n} |S_i|^r \gamma - \gamma| = O_p\left(\frac{1}{\sqrt{n}}\right)$. In addition, recall that

$$\hat{f}(U_i) = \frac{1}{(n-1)h} \sum_{j \neq i} K\left(\frac{U_j - U_i}{h}\right)$$

305
and

\[ \hat{f}(\hat{U}_i) = \frac{1}{(n-1)h} \sum_{j \neq i} K(\hat{U}_j - \hat{U}_i), \]

we have

\[
\max_{1 \leq i \leq n} |f(U_i) - \hat{f}(\hat{U}_i)| \leq \max_{1 \leq i \leq n} |f(U_i) - \hat{f}(U_i)| + \max_{1 \leq i \leq n} |\hat{f}(\hat{U}_i) - \hat{f}(U_i)| = O_p(\sqrt{\frac{L_n}{nh}}) + O_p\left(\frac{1}{\sqrt{nh^3}}\right).
\]

For the second one,

\[
|\hat{f}(\hat{U}_i)| \mathbb{1}\{U_i \in S_n^+\} \geq f(U_i) \left(1 - \frac{|\hat{f}(\hat{U}_i) - f(U_i)|}{f(U_i)}\right) \mathbb{1}\{U_i \in S_n^+\}
\]

\[
\geq f(U_i) \left(1 - O_p\left(\sqrt{\frac{L_n}{n^{1-2\rho-4H}}}\right)\right) \mathbb{1}\{U_i \in S_n^+\}
\]

\[
\geq n^{-\rho} L_n \mathbb{1}\{U_i \in S_n^+\},
\]

in which the first inequality is by triangle inequality, the second inequality is by Lemma C.0.13(2), and the last inequality is by Lemma C.0.13(2) and the fact that \(1 - 2\rho - 4H > 0\). \(\square\)
Bibliography


313


to_the_end
Biography

Yichong Zhang, the author of this dissertation, was born in February 24th, 1986. He obtained his Bachelor degree from Zhejiang University in June 2008 and his M.A. degree from Duke University in May 2011.