INFERENCE ON PARTIALLY IDENTIFIED PARAMETERS: WITH APPLICATIONS TO THE EVALUATION OF HETEROGENEOUS TREATMENT EFFECTS

By

Sang Soo Park

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Economics

August, 2008

Nashville, Tennessee

Approved:

Professor Yanqin Fan

Professor Kathryn Anderson

Professor James Foster

Professor Tong Li

Professor Bryan Shepherd

To my parents of whose lives I am proud

ACKNOWLEDGMENTS

I could not possibly thank all who provided help with this dissertation. First of all, I give special thanks to my advisor Professor Yanqin Fan. She was more than an advisor throughout my Ph.D. course. She was a mentor, helper, teacher (of course), and coauthor. Her guidance was essential for the completion of my dissertation. More importantly, I should say it was very enjoyable to have worked with her. Also, I would like to acknowledge the valuable suggestions and comments of my committee members: Professors Kathryn H. Anderson, James E. Foster, Tong Li, and Bryan Shepherd. There is no doubt that their suggestions and comments greatly enhanced the quality of my dissertation.

I cannot mention anyone else without thanking my colleagues, Mostafa Beshkar, Linda K. Carter, Robert G. Hammond, Ali Sina Ondar, Tommaso Tempesti, and Jisong Wu. They shared ideas with me, endeavored to solve problems together, and broadened the scope of my knowledge with so many discussions. The life in Calhoun 413C was a paragon of cooperative and constructive academic life. Of those, I especially thank Linda and Bob for having spent time on my writings.

There are several Korean friends whom I also want to thank. Professor Seung Jun Kwak advised me to apply for Vanderbilt. Dr. Ki Ho Lee offered me incredible encouragement. Dr. Jong Hun Kim was my mentor during my early years at Vanderbilt. Dr. Kyungwook Choi shared his knowledge on Econometrics. Jinsu Choi, Dongkuk Paeng, Seungwon Hu, Dr. Seungchan Yang, Dr. Jongyun Heo, and Dr. Seonghun Hong have shared insightful and useful discussions, which sharpened my intellectual capability.

I want to mention that the Noel dissertation fellowship, sponsored by Mr. Walter Noel, enabled me to concentrate on research without performing teaching assistant duties in the most critical semester. This facilitated the timely completion of my dissertation.

Finally, I want to thank my family; my parents of whose lives I am proud; my wife who sacrificed and supported my pursuit of a Ph.D.; my daughter who sacrificed her familiar life in Korea and agreed to come to the US to help her dad; and my two brothers who always support my decisions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

My dissertation focuses on the partial identification of distribution function of treatment effects. When individuals benefit differently from a given program (i.e., when treatment effects are believed to be heterogeneous), policy evaluation often requires knowledge of the distribution function of potential treatment effects. These cannot be point-identified unless information is available on the dependence structure between the potential outcomes with and without the treatment. At best, it can be partially identified by finding the upper and lower bounds. This dissertation develops a new approach to the partial identification and inference problem. Employing this approach for the evaluation of treatment effects expands the scope of the existing treatment effects literature toward a more explicit consideration of the heterogeneity of treatment effects.

The first essay presents statistical inference tools that are used in later chapters. It analyzes the inference problem for some partially identified parameters initiated and solved by Imbens and Manski (2004). Despite its beauty and simplicity, their solution has two flaws: that it assumes the superefficiency of the asymptotic distribution of the estimator for each bound; and that it is applicable only to one-dimensional parameters. Stoye (2007) solved the former problem by using the shrinkage estimators but left the second unsolved. On the other hand, the general inference tools for parameters defined by moment equalities and inequalities proposed by Andrews and Guggenberger (2007) are applicable to high dimensional parameters but their confidence sets are potentially conservative. In this essay, we exploit the duality between hypothesis testing and the construction of confidence sets to derive new confidence sets that are asymptotically uniformly valid under the same assumptions as in Stoye (2007) and Andrews and Guggenberger (2007). The confidence sets are generalizable to parameters defined by moment equalities and inequalities as in Andrews and Guggenberger (2007), and, in addition, are non-conservative.

The second essay considers treatment effects in the context of a randomized experiment. Contrary to existing literature on the treatment effects, whose interest is mainly the identification of the average treatment effects (ATE), the parameter of interest in the second essay is the distribution of treatment effects. The ATE delivers insufficient information on the treatment effects when treatment effects are heterogeneous, which is the case in many program evaluations (e.g., see Bitler, Gelbach, and Hoynes (1996)). We study the distribution of treatment effects since it is one of the measures that can further our understanding of heterogeneous treatment effects. Without imposing a particular dependence structure between the treated and control outcomes, however, the heterogeneity and unobservability of individual treatment effects preclude point-identification of the distribution of the treatment effects. At best, it is possible only to identify pointwise sharp bounds in which the true distribution lies.

In this essay, we examine partial identification of the distribution of treatment effects, propose a nonparametric estimator for each bound, and derive the asymptotic distribution of the estimator. Due to the asymptotic distribution's discontinuity on the model parameters, the confidence sets we proposed in the first essay are not directly applicable here. In the second essay, therefore, we develop only the procedure for inference on each bound and leave the inference on the true parameter for future work.

The discontinuity of the asymptotic distributions of the estimators for the bounds of the distribution of treatment effects has led our interest to the quantiles of the treatment effects. It should be recognized that the quantiles of treatment effects (Q_{TE}) are different from the quantile treatment effects (QTE) which are widely used in treatment effects literature. The QTE is only a special case of the Q_{TE} where the treated and control outcomes are assumed to exhibit the rank preservation property. However, this assumption is often too restrictive and, sometimes, even, unrealistic. In the third essay, we focus on the quantiles of treatment effects which allow any dependence structure. Here, we show the bounds for the quantiles of treatment effects, propose nonparametric estimators, and provide their asymptotic distribution. Since the limiting distribution of the estimators for the bounds is asymptotically normal, the inference methods developed in the first essay are applicable to the quantile of treatment effects.

We also propose new methods of constructing confidence intervals for the bounds and the true quantiles. Contrary to conventional methods which require the estimation of distribution density functions of the marginal distributions of the treated and control outcomes, our new methods do not need that information. Monte Carlo simulations with various marginal distributions show the new methods outperform other methods.

The fourth essay is an empirical study on treatment effects in Project STAR. Project STAR is a randomized experiment designed to investigate the effects of class size reduction (CSR) on students' performances. Based on data from this experiment, a sizable literature has already emerged to examine the positive effects of class size reduction (e.g., see Folger and Breda (1989), Nye, Hedges, and Konstantopoulos (1999), and Krueger and Whitmore (2001)). However, most work in the existing literature considers only the ATE, abstracting away from the possibility of heterogeneous treatment effects. A notable exception is Ding and Lehrer (2005), who reject the homogeneous treatment effects assumption and use the QTE to study heterogeneous treatment effects of Project STAR.

In this essay, I estimate the bounds for the Q_{TE} of CSR and examine the heterogeneity of the program effects. In addition to application of theories developed in previous essays, I also propose methodology that allows partial identification of the distribution of treatment effects conditional on pre-treatment outcomes. This methodology facilitates answering the question of "who benefits and who loses." The results confirm that: i) there existed heterogeneity in the CSR effects; ii) the pattern of heterogeneity in treatment effects might not be as simple as Ding and Lehrer (2005) asserted; iii) different subgroups might have different patterns of heterogeneity in treatment effects; iv) there could be students who were actually worse off due to the CSR.

CHAPTER II

CONFIDENCE SETS FOR SOME PARTIALLY IDENTIFIED PARAMETERS

Introduction

Partial identification of parameters of interest is common in many areas of economics, see Manski (2003) for a survey in microeconometrics, Chernozhukov, Hong, and Tamer (2007) (CHT henceforth) for an extensive list of examples in microeconomics, and Moon and Schorfheide (2007) for examples in macroeconomics. The distribution and quantile of the effects of a binary treatment studied in Chapter III and IV for randomized experiments and Fan and Wu (2007) for switching regimes models add to the already extensive list of partially identified parameters.

In the seminal paper of Imbens and Manski (2004) (IM henceforth), they proposed confidence intervals (CI) for interval identified parameters that are asymptotically uniformly valid under maintained assumptions. Since IM, numerous papers on inference for partially identified parameters have appeared in the literature. They can be classified into two groups; those based on re-sampling techniques such as subsampling and bootstrap; and those that do not reply on re-sampling. The former includes Bugni (2006), CHT, Romano and Shaikh (2005a,b) and the latter includes IM, Stoye (2007), Rosen (2005), Soares (2006), Beresteanu and Molinari (2006), and Andrews and Guggenberger (2007) (AG (2007) henceforth). More recently, Moon and Schorfheide (2007) present a Bayesian approach to this problem.

The simplicity of the CIs of Imbens and Manski (2004) and Stoye (2007) makes them appealing, but their dependence on the specific structure of interval identified parameters and the asymptotic normality of estimators of the lower and upper bounds on the true parameter makes them hard to generalize to parameters defined by general moment equalities/inequalities. In a series of papers, Andrews and Guggenberger (2005a,b,c, 2007, **AG hereafter**) developed several general methods of constructing uniform confidence sets (CS) in non-regular models. In AG (2007), they propose a simple plug-in asymptotic CS (PA-CS) for parameters defined by moment equalities/inequalities. Compared with the subsampling CS, AG (2007) showed that the PA-CS may be asymptotically conservative when there are restrictions on moment inequalities such that if one moment inequality holds as an equality, then another moment inequality can not be satisfied as an equality. A notable example of this is the interval identified parameter case unless the true parameter is point identified. In contrast, the CIs of IM and Stoye (2007) take into account such restriction and are not asymptotically conservative.

One contribution of the current chapter is to extend the CI of IM to parameters defined by general moment equalities/inequalities. To do this, we first re-examine the set-up of IM by using the general approach of constructing CSs by inverting a two-sided hypothesis test for the true parameter. We obtain an asymptotically uniformly valid, non-conservative CI by taking into account the restriction on the interval bounds and we show that it reduces to that of IM when there exists a super-efficient estimator of the length of the identified interval. We also show that the CI of Stoye (2007) can be obtained by inverting two one-sided tests for the true parameter. Unlike the CI of Stoye (2007), our CI shares the natural nesting property with that of IM, i.e., the CI with a larger nominal confidence level includes the CI with a smaller nominal confidence level. As a by-product, we note that our CI can be easily adapted to the case where estimators of the lower and upper bounds on the true parameter are not asymptotically normally distributed, provided their asymptotic distribution does not exhibit a discontinuity as a function of parameters of the model.

For interval identified parameters, the CI of Stoye (2007) and our new CI take into account the restriction on the interval bounds by estimating the length of the identified interval with a shrinkage estimator. To construct asymptotically non-conservative CSs for parameters defined by general moment equalities/inequalities, we use shrinkage estimators of the so-called slackness parameters, one for each moment inequality. The value of a slackness parameter reveals to what extent the correpsponding moment inequality is binding. For interval identified parameters, a weighted sum of the two slackness parameters is identical to the length of the identified interval and the use of shrinkage estimators of the slackness parameters plays the same role as the use of a shrinkage estimator of the length of the identified interval. Compared with existing CSs for parameters defined by moment equalities/inequalities, our CS is easy to implement; no re-sampling is required and no optimization is involved.

We carried out a simulation study on interval data and applied our new confidence interval, that of Stoye (2007), and the PA-CS of AG (2007) to three artificially created DGPs from the March 2000 wave of the Current Population Survey (CPS) data. The three DGPs represent respectively the point identified case, interval identified case with a small interval length, and interval identified case with a large interval length. Our general finding is that our new confidence interval and that of Stoye (2007) perform comparably, but the PA-CS of AG (2007) can over-cover especially when the sample size is large. Moreover, the simulation results support the theoretical finding of Stoye (2007) and the current chapter, i.e., it is essential to use the shrinkage estimator when the length of the identified interval is zero or small.

The rest of this chapter is organized as follows. In Section 2, we re-examine the case of interval identified parameters and construct a new CI for the true parameter by inverting a two-sided hypothesis test. In addition, we show that the CI of Stoye (2007) can be obtained by inverting two one-sided tests. In Section 3, we extend our new CI for interval identified parameters to a CS for parameters defined by general moment equalities/inequalities and show that it is asymptotically uniformly valid and non-conservative. Section 4 presents a simulation study and Section 5 concludes. Technical proofs are presented in Appendix A and some algebraic derivations are given in Appendices B and C.

Confidence Intervals for Interval Identified Parameters

Let $\theta_l \leq \theta_0 \leq \theta_u$, where $\theta_0 = \theta_0(P)$ is the object of interest which depends on a probability distribution P; P must lie in a set $\mathcal{P}$ that is characterized by ex ante constraints. The bounds θ_l, θ_u are identified, but θ_0 may not be. IM first introduced a uniform confidence interval (CI) for θ_0 under the assumption of asymptotic joint normality of $\hat{\theta}_l, \hat{\theta}_u$ and other assumptions, including super-efficiency of the estimator of $\Delta \equiv \theta_u - \theta_l$, where $\hat{\theta}_l, \hat{\theta}_u$ are consistent estimators of θ_l, θ_u respectively. Stoye (2007) proposed a CI that does not depend on the super-efficiency condition used in IM.

Useful examples of partial identification in some economic situations are illustrated below starting with the examples in IM. Other examples of interval identified parameters include the two-sided mean/interval data example, the quantile/distribution of the treatment effects in Chapter III and IV, and the correlation coefficient between the potential outcomes in a Gaussian switching regimes model (SRM) in Vijverberg (1993).

Example 1 (Two-Sided Mean/Interval Data). The parameter of interest is the population mean of a random variable Y, $E(Y)$. We do not observe the realizations of Y, but rather we observe the realizations of two random variables Y_L, Y_U such that $P(Y_L \leq Y \leq Y_U) = 1$. Let $\{Y_{Li}, Y_{Ui}\}_{i=1}^n$ be i.i.d. with the same distribution as $\{Y_L, Y_U\}$.

Let $\theta_l = E(Y_L)$ and $\theta_u = E(Y_U)$. Both θ_l and θ_u are point-identified from the sample information, but the parameter of interest $\theta_0 = E(Y)$ is interval identified unless $\theta_l = \theta_u$: $\theta_l \leq \theta_0 \leq \theta_u$. The estimators of the lower and upper bounds are given by $\hat{\theta}_l = n^{-1}\sum_{i=1}^{n} Y_{Li}$ and $\hat{\theta}_u = n^{-1}\sum_{i=1}^{n} Y_{Li}$.

Example 2 (Quantile of the Treatment effects). We consider a binary treatment and use Y_1 to denote the potential outcome from receiving treatment and Y_0 the outcome without treatment. Let $F_1(\cdot)$ and $F_0(\cdot)$ denote the distribution functions of Y_1 and Y_0 respectively. Let $\Delta = Y_1 - Y_0$ denote the treatment effects and $F_\Delta(\cdot)$ its distribution function. Given the marginals F_1 and F_0, sharp bounds on the quantile function of the treatment effects Δ can be found in Williamson and Downs (1990), see also Chapter III. Specifically, for $0 < p < 1$, let $\theta_0 = F_\Delta^{-1}(p)$,

$$\theta_l = \inf_{u\in[p,1]}[F_1^{-1}(u) - F_0^{-1}(u-p)], \text{ and } \theta_u = \sup_{u\in[0,p]}[F_1^{-1}(u) - F_0^{-1}(1+u-p)].$$

It is known that $\theta_l \leq \theta_0 \leq \theta_u$. With randomized data, F_1 and F_0 are identified and thus θ_l, θ_u are identified. Estimators of θ_l, θ_u can be constructed by replacing F_1 and F_0 with their consistent estimators such as the empirical distributions in the above expressions.

Example 3 (Correlation Between the Outcomes). Consider the following SRM:

$$\begin{aligned} Y_{1i} &= X_i'\beta_1 + U_{1i}, \\ Y_{0i} &= X_i'\beta_0 + U_{0i}, \\ D_i &= I_{\{W_i'\gamma+\epsilon_i>0\}}, \ i = 1, \ldots, n, \end{aligned} \tag{II.1}$$

where $\{X_i, W_i\}$ denote individual i's observed covariates and $\{U_{1i}, U_{0i}, \epsilon_i\}$ individual i's unobserved covariates. Here, D_i is the binary variable indicating participation of individual

i in the program or treatment; it takes the value 1 if individual i participates in the program and takes the value zero if she chooses not to participate in the program, Y_{1i} is the outcome of individual i we observe if she participates in the program, and Y_{0i} is her outcome if she chooses not to participate in the program. For individual i, we always observe the covariates $\{X_i, W_i\}$, but observe Y_{1i} if $D_i = 1$ and Y_{0i} if $D_i = 0$. The errors or unobserved covariates $\{U_{1i}, U_{0i}, \epsilon_i\}$ are assumed to be independent of the observed covariates $\{X_i, W_i\}$. We also assume the existence of an exclusion restriction, i.e., there exists at least one element of W_i which is not contained in X_i.

The textbook Gaussian model assumes that $\{U_{1i}, U_{0i}, \epsilon_i\}$ is trivariate normal:

$$\begin{pmatrix} U_{1i} \\ U_{0i} \\ \epsilon_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_0\rho_{10} & \sigma_l\rho_{1\epsilon} \\ \sigma_1\sigma_0\rho_{10} & \sigma_0^2 & \sigma_0\rho_{0\epsilon} \\ \sigma_l\rho_{1\epsilon} & \sigma_0\rho_{0\epsilon} & 1 \end{pmatrix}\right) \tag{II.2}$$

Based on the sample information alone, ρ_{10} is not identified. Using the fact that the covariance matrix of the errors is positive semi-definite, Vijverberg (1993) showed $\rho_L \leq \rho_{10} \leq \rho_U$, where

$$\rho_L = \rho_{1\epsilon}\rho_{0\epsilon} - \sqrt{(1-\rho_{1\epsilon}^2)(1-\rho_{0\epsilon}^2)},\ \rho_U = \rho_{1\epsilon}\rho_{0\epsilon} + \sqrt{(1-\rho_{1\epsilon}^2)(1-\rho_{0\epsilon}^2)}.$$

Note that ρ_L and ρ_U depend on the identified parameters only and hence are themselves identified, but ρ_{10} is only interval identified unless $\rho_L = \rho_U$. Estimators of ρ_L, ρ_U are straightforward to construct once the parameters $\rho_{1\epsilon}, \rho_{0\epsilon}$ are estimated by standard methods including maximum likelihood or the two-step approach of Heckman.

While Example 1 falls in the framework of parameters defined by moment inequalities, Examples 2 and 3 do not.

Review of IM and Stoye (2007)

IM proposed a CI for θ_0 as follows:

$$CI_{\text{IM}} \equiv \left[\hat{\theta}_l - \frac{c_\alpha \hat{\sigma}_l}{\sqrt{n}}, \hat{\theta}_u + \frac{c_\alpha \hat{\sigma}_u}{\sqrt{n}}\right],$$

where c_α solves

$$\Phi\left(c_\alpha + \frac{\sqrt{n}\hat{\Delta}}{\max\{\hat{\sigma}_l, \hat{\sigma}_u\}}\right) - \Phi(-c_\alpha) = 1 - \alpha. \tag{II.3}$$

in which $\hat{\Delta} = \hat{\theta}_u - \hat{\theta}_l$ and $\hat{\theta}_l, \hat{\theta}_u, \hat{\sigma}_l, \hat{\sigma}_u$ are defined in the following assumptions. These are the assumptions under which IM show the uniform validity of CI_{IM}.

Assumption IM (i) There are estimators $\hat{\theta}_l, \hat{\theta}_u$ that satisfy

$$\sqrt{n}\begin{pmatrix} \hat{\theta}_l - \theta_l \\ \hat{\theta}_u - \theta_u \end{pmatrix} \Rightarrow N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_l^2 & \rho\sigma_l\sigma_u \\ \rho\sigma_l\sigma_u & \sigma_u^2 \end{pmatrix}\right)$$

uniformly in $P \in \mathcal{P}$, and there are estimators $\left(\hat{\sigma}_l^2, \hat{\sigma}_u^2, \hat{\rho}\right)$ that converge to their population values uniformly in $P \in \mathcal{P}$.

(ii) For all $P \in \mathcal{P}$, $\underline{\sigma}^2 \leq \sigma_l^2, \sigma_u^2 \leq \overline{\sigma}^2$ for some positive and finite $\underline{\sigma}^2$ and $\overline{\sigma}^2$, and $\Delta \leq \overline{\Delta} < \infty$.

(iii) For all $\epsilon > 0$, there are $v > 0, K$, and N_0 such that $n \geq N_0$ implies that

$$\Pr\left(\sqrt{n}|\widehat{\Delta} - \Delta| > K\Delta^v\right) < \epsilon$$

uniformly in $P \in \mathcal{P}$.

Under Assumption IM (i)-(iii), IM showed that

$$\lim_{n\to\infty} \inf_{\theta\in\Theta} \inf_{P:\theta_0(P)=\theta} P\left(\theta_0 \in CI_{\text{IM}}\right) = 1 - \alpha,$$

i.e., CI_{IM} is asymptotically uniformly valid ($\lim_{n\to\infty}\inf_{\theta\in\Theta}\inf_{P:\theta_0(P)=\theta} P\left(\theta_0 \in CI_{\mathrm{IM}}\right) \geq 1-\alpha$); and non-conservative ($\lim_{n\to\infty}\inf_{\theta\in\Theta}\inf_{P:\theta_0(P)=\theta} P\left(\theta_0 \in CI_{\mathrm{IM}}\right) = 1-\alpha$).

Stoye (2007) pointed out that Assumption IM (iii) is a super-efficiency condition on the estimator $\widehat{\Delta}$ of the length of the identified interval and may be violated in important applications. In addition, Assumption IM (i)-(ii) and (iii) are mutually consistent for sequences of distributions P_n such that $\Delta_n \to 0$ only if $\sigma_l^2 - \sigma_u^2 \to 0$ and $\rho \to 1$ for all those sequences. To relax Assumption IM (iii), Stoye (2007) proposed the following CI for θ_0 and verified its asymptotic uniform validity and non-conservativeness under Assumption IM (i) and (ii) only:

$$CI_{\mathrm{S}} \equiv \begin{cases} \left[\hat{\theta}_l - \frac{c_l\hat{\sigma}_l}{\sqrt{n}}, \hat{\theta}_u + \frac{c_u\hat{\sigma}_u}{\sqrt{n}}\right] & \text{if } \hat{\theta}_l - \frac{c_l\hat{\sigma}_l}{\sqrt{n}} \leq \hat{\theta}_u + \frac{c_u\hat{\sigma}_u}{\sqrt{n}} \\ \varnothing & \text{otherwise} \end{cases},$$

where (c_l, c_u) minimize $(c_l\hat{\sigma}_l + c_u\hat{\sigma}_u)$ subject to the constraint that

$$\begin{aligned} \int_{-\infty}^{c_l} \Phi\left(\frac{\hat{\rho}}{\sqrt{1-\hat{\rho}^2}}z + \frac{c_u\hat{\sigma}_u + \sqrt{n}\Delta^*}{\hat{\sigma}_u\sqrt{1-\hat{\rho}^2}}\right) d\Phi(z) &\geq 1-\alpha, \\ \int_{-\infty}^{c_u} \Phi\left(\frac{\hat{\rho}}{\sqrt{1-\hat{\rho}^2}}z + \frac{c_l\hat{\sigma}_l + \sqrt{n}\Delta^*}{\hat{\sigma}_l\sqrt{1-\hat{\rho}^2}}\right) d\Phi(z) &\geq 1-\alpha, \end{aligned} \tag{II.4}$$

if $\hat{\rho} < 1$ and

$$\begin{aligned} \Phi(c_l) - \Phi\left(-\frac{c_u\hat{\sigma}_u + \sqrt{n}\Delta^*}{\hat{\sigma}_u}\right) &\geq 1-\alpha, \\ \Phi(c_u) - \Phi\left(-\frac{c_l\hat{\sigma}_l + \sqrt{n}\Delta^*}{\hat{\sigma}_l}\right) &\geq 1-\alpha, \end{aligned}$$

if $\hat{\rho} = 1$, in which Δ^* is a shrinkage estimator of Δ defined as

$$\Delta^* = \begin{cases} \hat{\Delta} \text{ if } \hat{\Delta} > b_n \\ 0 \text{ if otherwise,} \end{cases} \tag{II.5}$$

and b_n is some pre-assigned sequence such that $b_n \to 0$ and $b_n\sqrt{n} \to \infty$. As shown in Stoye

(2007), if Assumption IM (iii) holds, then CI_{S} reduces to that of IM (2004) except that CI_{S} uses Δ^* and CI_{IM} uses $\hat{\Delta}$.

New Confidence Interval for θ_0

The CIs of IM and Stoye (2007) are compositionally simple, but they rely heavily on the asymptotic normality of $\left(\hat{\theta}_l, \hat{\theta}_u\right)$, i.e., Assumption IM (i), and the specific structure of the identified set $[\theta_l, \theta_u]$ through the use of $\hat{\Delta}$ or Δ^*, see e.g., (II.3) and (II.4). As pointed out in Rosen (2005), Soares (2006), Pakes, Porter, Ho, and Ishii (2006) (PPHI henceforth), and AG (2007), many economic models imply moment equality/inequality constraints on parameters of interest and the identified set for these parameters may not be of the simple interval form.

In this subsection, we re-visit the issue of constructing CIs for interval identified parameter θ_0 by using the general approach of inverting a hypothesis test, aiming at understanding the roles played by the asymptotic normality of $\left(\hat{\theta}_l, \hat{\theta}_u\right)$ and the estimator of the length of the identified interval. By taking into account the interval structure of the identified set for θ_0, we establish an asymptotically non-conservative CI and show its uniform validity under Assumption IM (i) and (ii) only. Like Stoye (2007), we show that our CI reduces to the CI of IM when superefficiency holds. Unlike the CI of Stoye (2007), our CI shares the natural nesting property with that of IM, i.e., CIs with a larger nominal confidence level include CIs with a smaller nominal confidence level. More importantly, this approach allows us to generalize the CI of IM to some asymptotically non-normally distributed $\left(\hat{\theta}_l, \hat{\theta}_u\right)$ and parameters defined by moment equalities/inequalities.

We follow the notation in AG (2007). So, $\gamma_1 = (\gamma_{1l}, \gamma_{1u})$ with $\gamma_{1l} = (\theta - \theta_l)/\sigma_l$ and $\gamma_{1u} = (\theta_u - \theta)/\sigma_u$, $\gamma_2 = (\theta, \rho)$, γ_3 denotes the remaining parameters in P. The parameter space is

$$\Gamma = \left\{ \begin{array}{c} \gamma \equiv (\gamma_1, \gamma_2, \gamma_3) : \text{for some } (\theta, P) \in \mathcal{P}, \text{ where } \mathcal{P} \text{ is defined in Assumption IM (i) and (ii),} \\ \gamma_{1l} \geq 0, \gamma_{1u} \geq 0, \sigma_u \gamma_{1u} + \sigma_l \gamma_{1l} = \Delta, -1 \leq \rho \leq 1. \end{array} \right\}$$

Noting that

$$\theta_0 = \arg\min_{\theta} \left\{ \left(\frac{\theta_l - \theta}{\sigma_l} \right)_+^2 + \left(\frac{\theta_u - \theta}{\sigma_u} \right)_-^2 \right\},$$

where $(x)_- = \min\{x, 0\}$, $(x)_+ = \max\{x, 0\}$, we use the test statistic $T_n(\theta_0)$ defined below to construct CSs for θ_0:

$$T_n(\theta_0) = n \left(\frac{\hat{\theta}_l - \theta_0}{\hat{\sigma}_l} \right)_+^2 + n \left(\frac{\hat{\theta}_u - \theta_0}{\hat{\sigma}_u} \right)_-^2 . \tag{II.6}$$

A $1-\alpha$ CS for θ_0 is defined as

$$CS_n = \{\theta : T_n(\theta) \leq c_{1-\alpha}(\theta)\},$$

where $c_{1-\alpha}(\theta)$ is an appropriately chosen critical value to guarantee that CS_n has uniform asymptotic coverage rate of $1-\alpha$. As discussed in AG (2007), other test statistics can be used as well, but CSs based on them may not reduce to the CI of IM with super-efficiency.

Let $\{\gamma_{\omega_n,h} : n \geq 1\} \equiv \{(\gamma_{\omega_n,h,1}, \gamma_{\omega_n,h,2}, \gamma_{\omega_n,h,3}) : n \geq 1\}$ denote a sequence of parameters in Γ for which $\omega_n^{1/2} \gamma_{\omega_n,h,1} \to h_1 \equiv (h_l, h_u)$, $\gamma_{\omega_n,h,2} \to h_2 \equiv (h_\theta, h_\rho)$. Define

$$H = \left\{ (h_1, h_2) \in R_\infty^4 : \exists \text{ a subsequence } \{\omega_n\} \text{ of } \{n\} \text{ and a sequence } \{\gamma_{\omega_n,h} : n \geq 1\} \right\}.$$

Let $h = (h_1, h_2)$ and J_h denote the limiting distribution of $T_n(\theta_0)$ under $\{\gamma_{\omega_n,h}\}$. We show in Appendix A that J_h is the distribution function of the random variable $(Z_{l,h_\rho} - h_l)_+^2 +$

$\left(Z_{u,h_\rho} + h_u\right)_-^2$, where

$$\begin{pmatrix} Z_{l,h_\rho} \\ Z_{u,h_\rho} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & h_\rho \\ h_\rho & 1 \end{pmatrix}\right).$$

Since J_h depends on h_2 only through h_ρ, we use $cv_{1-\alpha}\left(h_l, h_u, h_\rho\right)$ to denote the $1-\alpha$ quantile of J_h. Likewise we denote J_h as $J_{(h_l,h_u,h_\rho)}$. We construct two CSs for θ_0 using J_h corresponding to different values of h. The first one defines the critical value $c_{1-\alpha}\left(\theta\right)$ in CS_n as $cv_{1-\alpha}\left(0,0,\hat{\rho}\right)$. This is the analog of PA-CS introduced in AG (2007) for parameters defined by moment equalities/inequalities. Specifically,

$$CI_{\text{AG}} = \left\{\theta : T_n(\theta) \leq cv_{1-\alpha}\left(0,0,\hat{\rho}\right)\right\}.$$

We show in Appendix B that CI_{AG} is in fact an interval, since $cv_{1-\alpha}\left(0,0,\hat{\rho}\right)$ does not depend on θ. Note that $h_l \geq 0$, $h_u \geq 0$, and J_h is stochastically decreasing in h_l, h_u. It follows that the PA-CS CI_{AG} is asymptotically uniformly valid, but it is in general conservative, as for any ρ, $(h_l, h_u, \rho) = (0, 0, \rho)$ may not belong to H unless $\theta_l = \theta_u$. This is because h_l, h_u satisfy $\sigma_u h_u + \sigma_l h_l = \lim\left(\sqrt{n}\Delta\right)$. In the special case where $\hat{\rho} = 1$, $J_{(0,0,1)}$ is $\chi^2_{[1]}$ and the PA-CS CI_{AG} reduces to the symmetric CI for the identification region $[\theta_l, \theta_u]$ first proposed in Horowitz and Manski (2000):

$$\left[\hat{\theta}_l - \frac{z_\alpha \hat{\sigma}_l}{\sqrt{n}}, \hat{\theta}_u + \frac{z_\alpha \hat{\sigma}_u}{\sqrt{n}}\right],$$

see also (2) in IM, where z_α is chosen such that

$$\Phi\left(z_\alpha\right) - \Phi\left(-z_\alpha\right) = 1 - \alpha.$$

An asymptotically non-conservative CI can be constructed by taking into account the restriction: $\sigma_u h_u + \sigma_l h_l = \lim(\sqrt{n}\Delta)$. Define

$$CI_{\text{FP}} = \left\{\theta : T_n(\theta) \leq c^*_{1-\alpha}(\hat{\rho})\right\},$$

where

$$c^*_{1-\alpha}(\hat{\rho}) = \max\left\{cv_{1-\alpha}\left(0, \frac{\sqrt{n}\Delta^*}{\hat{\sigma}_u}, \hat{\rho}\right), cv_{1-\alpha}\left(\frac{\sqrt{n}\Delta^*}{\hat{\sigma}_l}, 0, \hat{\rho}\right)\right\}, \tag{II.7}$$

in which Δ^* is the shrinkage estimator defined in (II.5).

Theorem 1 *Suppose Assumption IM (i) and (ii) hold and $0 < \alpha < 1/2$. Then CI_{FP} satisfies $\lim_{n\to\infty} \inf_{\theta\in\Theta} \inf_{P:\theta_0(P)=\theta} \Pr(\theta_0 \in CI_{FP}) = 1-\alpha$.*

Similar to CI_{AG}, CI_{FP} is an interval, as $c^*_{1-\alpha}(\hat{\rho})$ does not depend on θ. As shown in Appendix B, if $\rho = 1$, then

$$\begin{aligned} J_h(x) &\equiv J_{(h_l,h_u,\rho)}(x) \\ &= \Phi\left(h_l + \sqrt{x}\right) - \Phi\left(-h_u - \sqrt{x}\right). \end{aligned}$$

Hence $c^*_{1-\alpha}(1)$ satisfies[1]

$$\begin{aligned} \Phi\left(\sqrt{c^*_{1-\alpha}(1)}\right) - \Phi\left(-\frac{\sqrt{n}\Delta^*}{\hat{\sigma}_u} - \sqrt{c^*_{1-\alpha}(1)}\right) &\geq 1-\alpha, \\ \Phi\left(\frac{\sqrt{n}\Delta^*}{\hat{\sigma}_l} + \sqrt{c^*_{1-\alpha}(1)}\right) - \Phi\left(-\sqrt{c^*_{1-\alpha}(1)}\right) &\geq 1-\alpha, \end{aligned}$$

or equivalently

$$\Phi\left(\frac{\sqrt{n}\Delta^*}{\max\{\hat{\sigma}_l, \hat{\sigma}_u\}} + \sqrt{c^*_{1-\alpha}(1)}\right) - \Phi\left(-\sqrt{c^*_{1-\alpha}(1)}\right) = 1-\alpha. \tag{II.8}$$

It follows from (II.8) and the form of CI_{FP} established in Appendix C that with super-efficiency or $\hat{\rho} = 1$, CI_{FP} reduces to the uniform CI for θ_0 proposed in IM except that CI_{FP} uses Δ^*, while IM uses $\hat{\Delta}$. In this sense, CI_{FP} can be regarded as a natural extension of

[1] As explicitly stated in II.8, the critical values for IM in II.3 are comparable with $\sqrt{c^*_{1-\alpha}(1)}$ instead of $c^*_{1-\alpha}(1)$.

IM to the general case without super-efficiency condition.

Remark. (i) It is easy to see that CI_{FP} is nested; (ii) The asymptotic validity of CI_{FP} with $c^*_{1-\alpha}(\hat{\rho})$ defined in (II.7) does not depend on the asymptotic normality of $\left(\hat{\theta}_l, \hat{\theta}_u\right)$, as long as the asymptotic distribution of $\left(\hat{\theta}_l, \hat{\theta}_u\right)$ does not exhibit discontinuity as a function of parameters in the model; (iii) The distribution of the treatment effects in Chapter III provides an example of interval identified parameters for which the asymptotic distribution of estimators of the sharp bounds exhibits discontinuity as a function of parameters in the model. Park (2007a) is working on an extension of CI_{FP} to inference for the distribution of the treatment effects for randomized data.

CI of Stoye (2007) — Revisited

Instead of inverting a two-sided test, we can also invert two one-sided tests for H_0. For example, define

$$T_{nl}(\theta_0) = n\left(\frac{\hat{\theta}_l - \theta_0}{\hat{\sigma}_l}\right)^2_+ \text{ and } T_{nu}(\theta_0) = n\left(\frac{\hat{\theta}_u - \theta_0}{\hat{\sigma}_u}\right)^2_- .$$

Then a CI for θ_0 can be defined as

$$\begin{aligned} \overline{CI}_{\mathrm{S}} &= \{\theta : T_{nl}(\theta) \le c_l \cap T_{nu}(\theta) \le c_u\} \\ &= \left\{\theta : \hat{\theta}_l - \sqrt{c_l}\frac{\hat{\sigma}_l}{\sqrt{n}} \le \theta \le \hat{\theta}_u + \sqrt{c_u}\frac{\hat{\sigma}_u}{\sqrt{n}}\right\}, \end{aligned} \tag{II.9}$$

where c_l, c_u are chosen to guarantee the correct level of coverage.[2] (II.9) reveals that $\overline{CI}_{\mathrm{S}}$ is of the same form as the CI proposed by Stoye (2007). Note that under $\left\{\gamma_{\omega_n,h}\right\}$,

$$\begin{pmatrix} T_{nl}(\theta_0) \\ T_{nu}(\theta_0) \end{pmatrix} \Rightarrow \begin{pmatrix} \left(Z_{l,h_\rho} - h_l\right)^2_+ \\ \left(Z_{u,h_\rho} + h_u\right)^2_- \end{pmatrix}.$$

[2]We changed the definition of c_l and c_u in (II.4) to be consistent with other parts in the chapter. As a result, c_l and c_u in (II.4) are $\sqrt{c_l}$ and $\sqrt{c_u}$ here. We will use $\sqrt{c_l}$ and $\sqrt{c_u}$ hereafter.

We obtain

$$
\begin{aligned}
& \lim_{n\to\infty} \inf_{\theta\in\Theta} \inf_{P:\theta_0(P)=\theta} \Pr\left(\theta_0 \in \overline{CI}_{\mathrm{S}}\right) \\
= & \inf_{H} \Pr\left(Z_{l,h_\rho} \leq h_l + \sqrt{c_l} \cap Z_{u,h_\rho} \geq -h_u - \sqrt{c_u}\right) \\
= & \inf_{h_\rho} \min \left\{ \begin{array}{c} \Pr\left(Z_{l,h_\rho} \leq \sqrt{c_l} \cap Z_{u,h_\rho} \geq -\frac{\sqrt{n}\Delta}{\sigma_u} - \sqrt{c_u}\right), \\ \Pr\left(\frac{\sqrt{n}\Delta}{\sigma_l} + Z_{l,h_\rho} \leq \sqrt{c_l} \cap Z_{u,h_\rho} \geq -\sqrt{c_u}\right) \end{array} \right\} \\
= & \inf_{h_\rho} \min \left\{ \begin{array}{c} \Phi\left(\sqrt{c_u} + \frac{\sqrt{n}\Delta}{\sigma_u}\right) - \Phi\left(-\sqrt{c_l}, \sqrt{c_u} + \frac{\sqrt{n}\Delta}{\sigma_u}; h_\rho\right), \\ \Phi\left(\sqrt{c_u}\right) - \Phi\left(-\sqrt{c_l} - \frac{\sqrt{n}\Delta}{\widehat{\sigma}_l}, \sqrt{c_u}; h_\rho\right) \end{array} \right\} \qquad \text{(II.10)}
\end{aligned}
$$

where

$$
\Phi(x, y; \rho) = \int_{-\infty}^{y} \int_{-\infty}^{x} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}\left(\frac{s^2 - 2\rho st + t^2}{1-\rho^2}\right)\right) ds dt.
$$

The second equality follows from concavity of $\Pr\left(Z_{l,h_\rho} \leq h_l + \sqrt{c_l} \cap Z_{u,h_\rho} \geq -h_u - \sqrt{c_u}\right)$ expressed as a function of h_l (Stoye 2007).

To determine c_l and c_u, we minimize the length of the $\overline{CI}_{\mathrm{S}} : \hat{\sigma}_u\sqrt{c_u} + \hat{\sigma}_l\sqrt{c_l} + \hat{\Delta}$ such that

$$
\begin{aligned}
& \min \left\{ \begin{array}{c} \Pr\left(Z_{l,\hat{\rho}} \leq \sqrt{c_l} \cap Z_{u,\hat{\rho}} \geq -\frac{\sqrt{n}\Delta^*}{\widehat{\sigma}_u} - \sqrt{c_u}\right), \\ \Pr\left(\frac{\sqrt{n}\Delta^*}{\widehat{\sigma}_l} + Z_{l,\hat{\rho}} \leq \sqrt{c_l} \cap Z_{u,\hat{\rho}} \geq -\sqrt{c_u}\right) \end{array} \right\} \\
= & \min \left\{ \begin{array}{c} \Phi\left(\sqrt{c_u} + \frac{\sqrt{n}\Delta^*}{\widehat{\sigma}_u}\right) - \Phi\left(-\sqrt{c_l}, \sqrt{c_u} + \frac{\sqrt{n}\Delta^*}{\widehat{\sigma}_u}; \hat{\rho}\right), \\ \Phi\left(\sqrt{c_u}\right) - \Phi\left(-\sqrt{c_l} - \frac{\sqrt{n}\Delta^*}{\widehat{\sigma}_l}, \sqrt{c_u}; \hat{\rho}\right) \end{array} \right\} \\
= & 1 - \alpha.
\end{aligned}
$$

It can be easily shown that this leads to the same CI as that of Stoye (2007).

Parameters Defined by Moment Equalities/Inequalities

We follow the notation of AG (2007). Suppose there exists a true value θ_0 that satisfies the moment conditions:

$$
\begin{aligned}
Em_j\left(W_i,\theta_0\right) &\geq 0 \text{ for } j=1,...,p \text{ and} \\
Em_j\left(W_i,\theta_0\right) &= 0 \text{ for } j=p+1,...,p+v,
\end{aligned} \tag{II.11}
$$

where $\{m_j(\cdot,\theta): j=1,...,p+v\}$ are known real-valued moment functions and $\{W_i: i\geq 1\}$ are observed i.i.d. random vectors[3] with joint distribution P. The true value θ_0 is not necessarily point identified, but the moment equalities/inequalities in (II.11) restrict the set of values of θ_0, referred to as the identified set of θ_0. In many economic/econometric models, the parameters of interest are defined by a finite number of moment equalities/inequalities in (II.11). One widely studied example of partially identified models in microeconometric literature is an entry game, see Bresnahan and Reis (1991), Berry (1992), Tamer (2003), and Ciliberto and Tamer (2004). In the simple version with only two players, depending on the entry decision of the second firm, Firm 1 either does not enter market, or operates as monopolist, or operates as duopolist. Assuming that the outcome of the entry game in each market is a pure strategy Nash equilibrium, it is straightforward to show that the Nash equilibrium is unique, except when both firms are profitable as monopolist but not as duopolist. In the latter case, the model is silent about which firm actually enters the market. As a result, it only delivers bounds for the probability of observing a particular monopoly.

Example 5 (Simultaneous Entry Game) Let Y_j be the player j's entry decision for $j=1,2$. $Y_j=1$ if the stochastic payoff function $\pi_j\left(Y_j,Y_{-j}\right)>0$; 0 otherwise. Let's assume a simple linear payoff function, that is, $\pi_j\left(Y_j,Y_{-j}\right)=X_j\beta_j-d_jY_{-j}+v_j$, $E\left[v_j|X_j,X_{-j}\right]=0$, and $d_j>0$. Then, because there exist multiple equilib-

[3] The i.i.d. assumption is made for ease of exposition. This can be relaxed, see AG (2007).

ria, $E\left[Y_1\left(1-Y_0\right)|X_1,X_2\right]=P\left(Y_1=1,Y_0=0|X_1,X_2\right)$ is partially identified i.e.

$$P_{(1,0)L}\leq P\left(Y_1=1,Y_0=0|X_1,X_2\right)\leq P_{(1,0)U}$$

where

$$\begin{aligned}
P_{(1,0)L} &= P\left(v_1>-X_1\beta_1+d_1,v_2\leq -X_2\beta_2+d_2\right)\\
&\quad +P\left(-X_1\beta_1<v_1\leq -X_1\beta_1+d_1,v_2\leq -X_2\beta_2\right),\\
P_{(1,0)U} &= P\left(v_1>-X_1\beta_1,v_2\leq -X_2\beta_2+d_2\right).
\end{aligned}$$

Similar bounds can be construct for $E\left[Y_1\left(1-Y_0\right)|X_1,X_2\right]=P\left(Y_1=0,Y_0=1|X_1,X_2\right).$

Another example is that of regression models with interval outcomes in Manski and Tamer (2002) . Additional examples can be found in the references in the Introduction.

Example 6 (Regression Models with Interval Outcomes) Suppose a regressor vector X_i is available and the conditional mean of unobserved Y_i is modeled using the linear function $X_i'\theta$. It is known that $P\left(Y_{Li}\leq Y_i\leq Y_{Ui}\right)=1.$ The parameter θ satisfies

$$E\left[Y_{Li}|X_i\right]\leq X_i'\theta\leq E\left[Y_{Ui}|X_i\right].$$

These conditional restrictions imply the inequalities

$$E\left[Y_{Li}Z_i\right]\leq \theta' E\left[X_iZ_i\right]\leq E\left[Y_{Ui}Z_i\right], \tag{II.12}$$

where Z_i is a vector of positive transformations of X_i, see CHT (2007). Let Z_i be of dimension q. This falls in the moment inequality framework of (II.11) with $p=2q,v=0$, see also CHT (2007), AG (2007), and Beresteanu and Molinari (2006).

In general, the identified set for θ_0 defined in (II.11) does not have a simple interval structure, preventing CI_{FP} and CI_{S} from being directly applicable. The purpose of this section is to extend CI_{FP} to θ_0 in (II.11) and clarify its relation to existing non-resampling based CSs in Rosen (2005), Soares (2006), PPHI (2006), and AG (2007).

Let

$$m(W_i,\theta) = (m_1(W_i,\theta), ..., m_k(W_i,\theta)),$$

where $k = p+v$. We make the same assumptions as AG (2007) and refer the reader to their paper for details. Define $\gamma_1 = (\gamma_{1,1}, ..., \gamma_{1,p})' \in R_+^p$ by writing the moment inequalities in (II.11) as moment equalities:

$$\sigma_j^{-1}(\theta) E m_j(W_i,\theta) - \gamma_{1,j} = 0 \text{ for } j = 1, ..., p,$$

where $\sigma_j^2(\theta) = Var(m_j(W_i,\theta))$. Moon and Schorfheide (2007) refer parameters $\gamma_{1,j}$, $j = 1, .., p$ to as the slackness parameters. Let

$$T_n(\theta) = n\sum_{j=1}^{p}\left[\frac{\overline{m}_{n,j}(\theta)}{\hat{\sigma}_{n,j}(\theta)}\right]_-^2 + n\sum_{j=p+1}^{p+v}\left[\frac{\overline{m}_{n,j}(\theta)}{\hat{\sigma}_{n,j}(\theta)}\right]^2,$$

where $\hat{\sigma}_{n,j}^2(\theta)$ is a consistent estimator of $\sigma_j^2(\theta)$. Let $\Omega = \Omega(\theta) = Corr(m(W_i,\theta))$.

Let $\gamma_2 = (\gamma_{2,1}, \gamma_{2,2}) = (\theta, vech_*(\Omega))$, where $vech_*(\Omega)$ denotes the vector of elements of Ω that lie below the main diagonal, and γ_3 the remaining parameters in the model. AG (2007) showed that under the local sequence $\{\gamma_{\omega_n,h}\}$,

$$T_n(\theta) \Longrightarrow J_h \equiv \sum_{j=1}^{p}\left[Z_{h_{2,2},j} + h_1\right]_-^2 + \sum_{j=p+1}^{p+v}\left[Z_{h_{2,2},j}\right]^2,$$

where $h = (h_1, h_2)$ in which $h_1 = \lim\left(\omega_n^{1/2}\gamma_{\omega_n,h,1}\right)$ and $h_2 \equiv (h_{2,1}, h_{2,2}) = \lim\left(\omega_n^{1/2}\gamma_{\omega_n,h,1}\right)$, $Z_{h_{2,2}} = \left(Z_{h_{2,2},1}, ..., Z_{h_{2,2},k}\right)' \sim N\left(0_k, \Omega_{h_{2,2}}\right)$ and $\Omega_{h_{2,2}}$ can be consistently estimated by

$$\hat{\Omega}_n(\theta) = \hat{D}_n^{-1/2}(\theta)\hat{\Sigma}_n(\theta)\hat{D}_n^{-1/2}(\theta)$$

with $\hat{D}_n(\theta) = Diag\left(\hat{\Sigma}_n(\theta)\right)$ and

$$\hat{\Sigma}_n(\theta) = n^{-1} \sum_{i=1}^{n} \left(m(W_i, \theta) - \overline{m}_n(\theta)\right) \left(m(W_i, \theta) - \overline{m}_n(\theta)\right)'.$$

Let $cv_{1-\alpha}(h_1, h_2)$ denote the $1-\alpha$ quantile of J_h. Note that two types of parameters appear in J_h: h_1 and $h_{2,2}$ or $\Omega_{h_{2,2}}$. To ease the exposition, we rewrite $cv_{1-\alpha}(h_1, h_2)$ as a function of h_1 and $\Omega_{h_{2,2}}$: $cv_{1-\alpha}\left(h_1, \Omega_{h_{2,2}}\right)$. Although $\Omega_{h_{2,2}}$ can be consistently estimated, h_1 can not. To circumvent this problem, AG (2007) proposed a PA-CS for θ_0 by using the critical value $cv_{1-\alpha}\left(0, \hat{\Omega}_n(\theta)\right)$. They show that the PA-CS is not asymptotically conservative provided there are no restrictions on the moment inequalities such that satisfaction of one inequality as an equality implies violation of another. But as they noted, such restrictions do arise in some examples, including the two-sided mean example and regression models with interval outcome data. In these examples, the vector of slackness parameters γ_1 is restricted to be in a subset of R_+^p. For example, for the two-sided mean or interval identified parameters, $\gamma_1 \in \{\gamma_{1l} \geq 0, \gamma_{1u} \geq 0, \sigma_u \gamma_{1u} + \sigma_l \gamma_{1l} = \Delta\} \subset R_+^2$ unless $\Delta = 0$. Provided θ_0 is not point identified, the restriction: $\sigma_u \gamma_{1u} + \sigma_l \gamma_{1l} = \Delta$, implies that if one inequality is satisfied as an equality, e.g., $\gamma_{1l} = 0$, then the other inequality can not be satisfied as an equality, as $\gamma_{1u} = \Delta/\sigma_u > 0$. By taking into account this specific structure or restriction on the moment inequalities, the CI we constructed for interval identified parameters are not asymptotically conservative. However, it does not allow for a straightforward generalization to the case characterized by general moment equalities/inequalities, as there is no such simple characterization of restrictions of this type. Instead we propose the following remedy: for $j = 1, ..., p$, we define

$$\gamma_{1,j}^*(\theta) = \begin{cases} \frac{\overline{m}_{n,j}(\theta)}{\hat{\sigma}_j(\theta)} & \text{if } \overline{m}_{n,j}(\theta) > b_n \\ 0 & \text{otherwise} \end{cases}$$

Let $\gamma_1^*(\theta) = \left(\gamma_{1,1}^*(\theta), ..., \gamma_{1,p}^*(\theta)\right)$ and define

$$CS_{\mathrm{MC}} = \left\{\theta : T_n(\theta) \leq cv_{1-\alpha}\left(\sqrt{n}\gamma_1^*(\theta), \hat{\Omega}_n(\theta)\right)\right\},$$

Theorem 2 *Under the same assumptions as Theorem 2 (a) of AG (2007), we have*

$$\lim_{n\to\infty} \inf_{\theta\in\Theta} \inf_{P:\theta_0(P)=\theta} \Pr(\theta_0 \in CS_{MC}) = 1 - \alpha.$$

It is interesting to observe that the CSs of Rosen (2005), Soares (2006), PPHI (2006), and the PA-CS of AG (2007) are all[4] based on $cv_{1-\alpha}\left(h_1, \hat{\Omega}_n(\theta)\right)$ except that they use different values of h_1: The CS of PPHI (2006) and the PA-CS of AG (2007) use $cv_{1-\alpha}\left(0, \hat{\Omega}_n(\theta)\right)$ and are asymptotically conservative when there are restrictions on the moment inequalities such that satisfaction of one inequality as an equality implies violation of another; Rosen (2005) and Soares (2006) use $cv_{1-\alpha}\left(0, .., 0, \infty, .., \infty, \hat{\Omega}_n(\theta)\right)$ with p^* zeros, where p^* is an upper bound on the number of binding inequality constraints in Rosen (2006) and p^* is the number of binding moment inequalities chosen via some moment selection criterion in Soares (2006). It is thus expected that the CS of Soares (2006) is less conservative than those of Rosen (2005), PPHI (2006), and the PA-CS of AG (2007). However, as Soares (2006) pointed out, this procedure may be compositionally intensive depending on the dimension of θ.

Interval-Identified Parameters. Instead of estimating $\Delta = \theta_u - \theta_l$ by the shrinkage estimator Δ^*, we estimate γ_{1l} and γ_{1u} by shrinkage:

$$\gamma_{1l}^* = \begin{cases} \frac{\theta-\hat{\theta}_l}{\hat{\sigma}_l} \text{ if } \theta - \hat{\theta}_l > b_n \\ 0 \quad \text{otherwise} \end{cases}, \ \gamma_{1u}^* = \begin{cases} \frac{\hat{\theta}_u-\theta}{\hat{\sigma}_u} \text{ if } \hat{\theta}_u - \theta > b_n \\ 0 \quad \text{otherwise} \end{cases}.$$

An alternative CS for θ_0 can be defined as follows:

$$CS_{\mathrm{IP}} = \left\{\theta : T_n(\theta) \leq cv_{1-\alpha}\left(\sqrt{n}\gamma_{1l}^*, \sqrt{n}\gamma_{1u}^*, \hat{\rho}\right)\right\}.$$

[4]Rosen (2005) uses a different test statistic from $T_n(\theta)$.

Note that the use of shrinkage estimators γ_{1l}^* and γ_{1u}^* in CS_{IP} automatically takes into account the restriction on the moment inequalities. To see this, suppose $\gamma_{1l} = 0$ so that $\theta = \theta_l$. This implies $\gamma_{1u} = \Delta > 0$ unless $\Delta = 0$. For large enough samples, $\theta - \hat{\theta}_l$ would be smaller than b_n and thus, $\gamma_{1l}^* = 0$. In contrast, γ_{1u}^* would approach Δ/σ_u. At the boundaries, the two CSs: CI_{FP} and CS_{IP} behave similarly.

Regression Models with Interval Outcomes. In addition to CS_{MC}, if $q = 1$, we can also extend CI_{FP} to θ_0. Let $W_i = (Y_{Li}, Y_{Ui}, X_i, Z_i)$,

$$m_1 (W_i, \theta) = \theta' [X_i Z_i] - Y_{Li} Z_i, \; m_2 (W_i, \theta) = Y_{Ui} Z_i - \theta' [X_i Z_i] .$$

Let

$$\begin{pmatrix} Z_{1,\rho} \\ Z_{u,\rho} \end{pmatrix} \Rightarrow N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_l^2 (\theta) & \sigma_l (\theta) \sigma_u (\theta) \rho (\theta) \\ \sigma_l (\theta) \sigma_u (\theta) \rho (\theta) & \sigma_u^2 (\theta) \end{pmatrix} \right)$$

and J_h denote the distribution function of the random variable $(Z_{l,\rho} - h_l)_+^2 + (Z_{u,\rho} + h_u)_-^2$ with $\rho = \rho(\theta)$. Note that $\Delta \equiv m_u (\theta) - m_l (\theta) = E [Y_{Ui} Z_i] - E [Y_{Li} Z_i]$ is point identified and can be consistently estimated by

$$\hat{\Delta} = \frac{1}{n} \sum_{i=1}^{n} (Y_{Ui} - Y_{Li}) Z_i.$$

This can be taken into account to construct a CS for θ_0 that is not asymptotically conservative. Let $cv_{1-\alpha} (h)$ denote the $1 - \alpha$ quantile of J_h. Note that the CS in AG (2007) uses the critical value $cv_{1-\alpha} (0, 0, \hat{\rho} (\theta))$, where

$$\hat{\rho} (\theta) = \frac{n^{-1} \sum_{i=1}^{n} [m_{li} (\theta) - \overline{m}_l (\theta)] [m_{ui} (\theta) - \overline{m}_u (\theta)]}{\hat{\sigma}_l (\theta) \hat{\sigma}_u (\theta)}.$$

We propose to use:

$$c_{1-\alpha}(\theta) = \sup_{0 \le h_l \le \frac{\sqrt{n}\Delta^*}{\hat{\sigma}_l(\theta)}} cv_{1-\alpha}\left(h_l, \frac{\sqrt{n}\Delta^* - \hat{\sigma}_l(\theta) h_l}{\hat{\sigma}_u(\theta)}, \hat{\rho}(\theta)\right), \tag{II.13}$$

in which Δ^* is a shrinkage estimator of Δ defined as

$$\Delta^* = \begin{cases} \hat{\Delta} \text{ if } \hat{\Delta} > b_n \\ 0 \text{ otherwise} \end{cases}.$$

Numerical Studies

In this section, we first present a numerical comparison of the critical values of five CIs at 0.95 nominal level: CI_{FP}, CI_{S}, CI_{AG}, and CI_{IM}, and then present some results from a small-scale simulation study on the finite sample performance of CI_{FP}, CI_{S}, and CI_{AG}.

Computation and Comparison of Critical Values

We recall that CI_{FP} uses $c^*_{1-\alpha}(\rho)$ in (II.7):

$$c^*_{1-\alpha}(\hat{\rho}) = \max\left\{cv_{1-\alpha}\left(0, \frac{\sqrt{n}\Delta^*}{\hat{\sigma}_u}, \hat{\rho}\right), cv_{1-\alpha}\left(\frac{\sqrt{n}\Delta^*}{\hat{\sigma}_l}, 0, \hat{\rho}\right)\right\},$$

where $cv_{1-\alpha}(h_l, h_u, \rho)$ is the $1-\alpha$ quantile of J_h for a given $h = (h_l, h_u, \rho)$ and J_h is the distribution function of the random variable, $(Z_{l,\rho} - h_l)^2_+ + (Z_{u,\rho} + h_u)^2_-$.

We first show that

$$c^*_{1-\alpha}(\hat{\rho}) = \begin{cases} cv_{1-\alpha}\left(\frac{\sqrt{n}\Delta^*}{\hat{\sigma}_l}, 0, \hat{\rho}\right) \text{ if } \hat{\sigma}_l \ge \hat{\sigma}_u \\ cv_{1-\alpha}\left(0, \frac{\sqrt{n}\Delta^*}{\hat{\sigma}_u}, \hat{\rho}\right) \text{ if } \hat{\sigma}_l < \hat{\sigma}_u \end{cases}. \tag{II.14}$$

From the symmetry of the joint distribution of $(Z_{l,\rho}, Z_{u,\rho})$, it follows that the random variables $(Z_{l,\rho})^2_+ + \left(Z_{u,\rho} + \frac{\sqrt{n}\Delta}{\sigma_u}\right)^2_-$ and $\left(Z_{l,\rho} - \frac{\sqrt{n}\Delta}{\sigma_u}\right)^2_+ + (Z_{u,\rho})^2_-$ have the same distribution function. But

$$\begin{aligned}
&\left\{\left(Z_{l,\rho}-\frac{\sqrt{n}\Delta}{\sigma_l}\right)_+^2+(Z_{u,\rho})_-^2\right\}-\left\{\left(Z_{l,\rho}-\frac{\sqrt{n}\Delta}{\sigma_u}\right)_+^2+(Z_{u,\rho})_-^2\right\}\\
=&\ \left(Z_{l,\rho}-\frac{\sqrt{n}\Delta}{\sigma_l}\right)_+^2-\left(Z_{l,\rho}-\frac{\sqrt{n}\Delta}{\sigma_u}\right)_+^2\\
\geq&\ 0 \text{ a.s. if } \sigma_l\geq\sigma_u;\ \leq 0 \text{ a.s. if } \sigma_l<\sigma_u,
\end{aligned}$$

implying (II.14).

So to compute $c^*_{1-\alpha}(\rho)$, we just need to compute $cv_{1-\alpha}\left(\frac{\sqrt{n}\Delta^*}{\hat{\sigma}_l},0,\hat{\rho}\right)$ or $cv_{1-\alpha}\left(0,\frac{\sqrt{n}\Delta^*}{\hat{\sigma}_u},\hat{\rho}\right)$ depending on which of $\hat{\sigma}_l,\hat{\sigma}_u$ is larger. One method for computing $cv_{1-\alpha}(h)$ for a given h is by simulation. Alternatively, one can invert J_h numerically. In Appendix B, we show that for $|\rho|<1$,

$$\begin{aligned}
J_h(x) &\equiv J_{(h_l,h_u,\rho)}(x)\\
&= \Phi\left(h_l+\sqrt{x}\right)-\int_{-\infty}^{h_l+\sqrt{x}}\Phi\left(-\frac{\rho z+h_u+\sqrt{x-(z-h_l)_+^2}}{\sqrt{1-\rho^2}}\right)d\Phi(z);
\end{aligned}$$

If $\rho=1$, then

$$J_h(x)=\Phi\left(h_l+\sqrt{x}\right)-\Phi\left(-h_u-\sqrt{x}\right);$$

Let $h_{\max}=\max\{h_l,h_u\}$ and $h_{\min}=\min\{h_l,h_u\}$. If $\rho=-1$, then

$$J_h(x)=\begin{cases}\Phi\left(h_{\min}+\sqrt{x}\right) \text{ if } x\leq(h_{\max}-h_{\min})^2\\ \Phi\left(\frac{h_{\max}+h_{\min}+\sqrt{2x-(h_{\max}-h_{\min})^2}}{2}\right) \text{ if } (h_{\max}-h_{\min})^2<x\end{cases}.$$

For any fixed x, the value of $J_h(x)$ can be computed numerically using the above expressions. We have written a Gauss program for computing $c^*_{1-\alpha}(\hat{\rho})$ which is available upon request.

The CIs: CI_{AG} and CI_{HM} are respectively based on $cv_{1-\alpha}(0,0,\rho)$ and $\sqrt{cv_{1-\alpha}(0,0,1)}$. In Figure 1, we plotted $\sqrt{cv_{0.95}(0,0,\rho)}$ against $\rho\in[-1,1]$. We note that $\sqrt{cv_{0.95}(0,0,\rho)}$ decreases as ρ increases and approaches to $\Phi^{-1}(1-\alpha/2)=1.96$ as $\rho\to1$. But for small values of ρ, $cv_{1-\alpha}(0,0,\rho)$ can be much larger than $cv_{1-\alpha}(0,0,1)$.

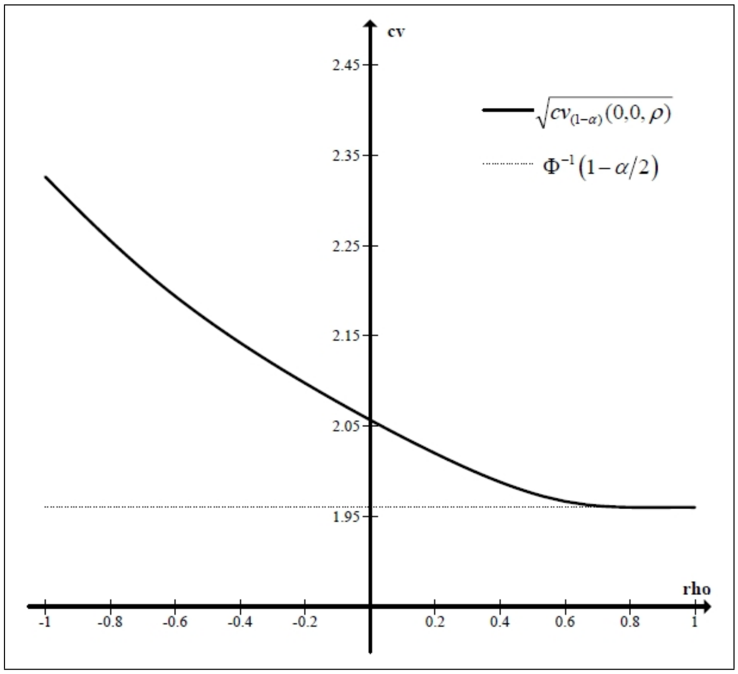


Figure 1. $\sqrt{cv_{0.95}(0,0,\rho)}$ and $\Phi^{-1}(0.975)$

In Figure 2, we plotted the critical values in CI_{FP}, CI_{S}, and CI_{IM} against $\sqrt{n}\Delta/\max\{\sigma_l,\sigma_u\}$ for $\rho = -0.4, 0, 0.4, 1$.

The critical values for CI_{FP} and CI_{IM} depend on σ_l, σ_u through $\sqrt{n}\Delta/\max\{\sigma_l,\sigma_u\}$ only. But the critical value of CI_{S} also depends on the values of σ_l, σ_u. We chose two sets of values: $(\sigma_l^2, \sigma_u^2) = (2,2)$ and $(\sigma_l^2, \sigma_u^2) = (1,2)$. When $\sigma_l^2 = \sigma_u^2$, Stoye's lower and upper critical values are the same. They are denoted as Stoye. When $\sigma_l^2 \neq \sigma_u^2$, they differ and are denoted as StoyeL and StoyeU respectively. In the graphs, StoyeL > StoyeU for all of the settings.

Several interesting conclusions can be made based on Figure 2. First, when $\sqrt{n}\Delta/\max\{\sigma_l,\sigma_u\} > 2.5$, all the critical values become almost identical to $\Phi^{-1}(1-\alpha) = 1.645$. Second, when $\sqrt{n}\Delta/\max\{\sigma_l,\sigma_u\}$ is small, the critical values for different CIs differ and the difference becomes larger as ρ approaches to -1. Third, when ρ is positive and

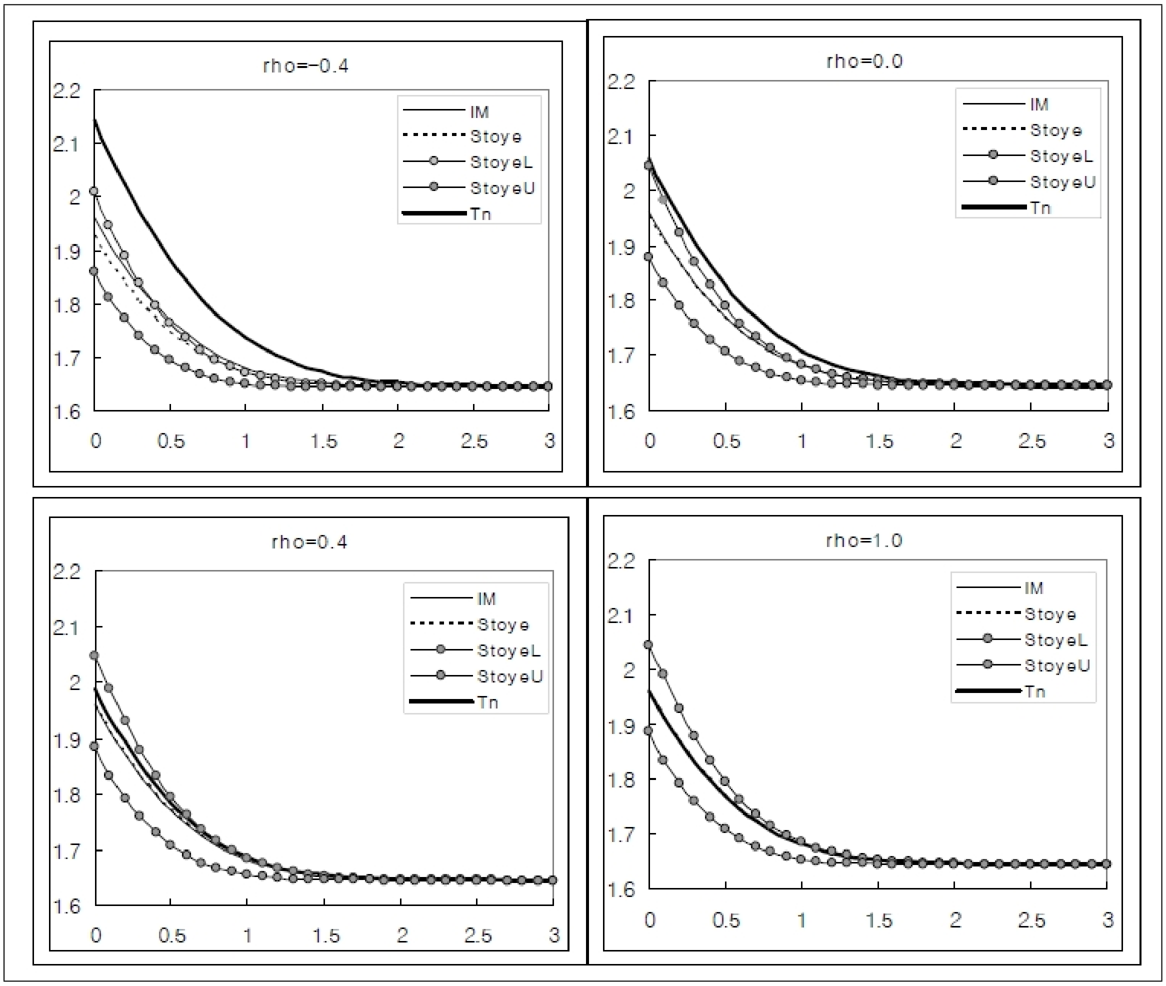


Figure 2. Comparison of Critical Values

$\sigma_l = \sigma_u$, the critical values of CI_{IM} and CI_{S} are numerically indistinguishable. Lastly, when $\rho = 1$, the critical values of CI_{FP} and CI_{IM} coincide and they coincide with that of CI_{S} if $\sigma_l = \sigma_u$. But if $\sigma_l \neq \sigma_u$, the critical values of CI_{S} differ from that of CI_{FP} or CI_{IM}.

Simulation: Population Mean with Interval Data

We apply CI_{FP}, CI_{S}, and CI_{AG} to the example of two-sided mean or interval data. Like CHT (2004) and Beresteanu and Molinari (2006), we use the March 2000 wave of the Current Population Survey (CPS) data. The variable Y is the logarithm of *wages* *and salaries* of white men ages 20 to 50 only. The 'population' of study consists of 13290 observations summarized in Table 1.

Table 1. Summary Statistics of DGP1: CPS Data

Variable	# of Values	Mean	Std Dev	Min	Max
$\exp(Y)$ (*wages and salaries*, in \$)	13290	66943.2	52465.0	1	513472
Y	13290	4.539	0.985	0	5.711

In the simulation, the 'population' or DGP consists of population values of the lower bound Y_L and the corresponding values of the upper bound Y_U. From this DGP, we draw random samples of sizes $n = 500, 1000, 2000, 8000$ respectively denoted as $\{Y_{Li}, Y_{Ui}\}_{i=1}^n$. The estimators of the lower and upper bounds are given by $\hat{\theta}_l = n^{-1} \sum_i Y_{Li}$ and $\hat{\theta}_u = n^{-1} \sum_i Y_{Li}$.

We considered three DGPs designed to shed light on the performance of CI_{FP}, CI_{S}, and CI_{AG} in three typical cases: point-identified case, interval identified case with a small Δ, and interval identified case with a large Δ. For point identified case, the DGP (DGP1) is the CPS data set, from which we draw two types of random samples $\{Y_{Li}, Y_{Ui}\}_{i=1}^n$; one with $Y_{Li} = Y_{Ui} = Y_i$ for $i = 1, ..., n$ and the other with $\{Y_{Li}\}_{i=1}^n, \{Y_{Ui}\}_{i=1}^n$ being independent. For interval identified case with small Δ, the DGP (DGP2) consists of the logarithms of the bracketed *wages and salaries* data in CHT (2004) and Beresteanu and Molinari (2006). There are 16 brackets: the values of Y_L and Y_U are the logarithms of the bracketed *wages and salaries*. These brackets are (written in thousand \$): $[0.001, 5], [5, 7.5], [7.5, 10], [10, 12.5], [12.5, 15], [15, 20], [20, 25], [25, 30],$

$[30, 35], [35, 40], [40, 50], [50, 60], [60, 75], [75, 100], [100, 150], [150, 100000]$ .For large Δ, we combined the first eight brackets into one: $[0.001, 30]$ and the last eight into the other one: $[30, 100000]$ and the DGP (DGP3) consists of the logarithms of the two bracketed *wages and salaries*. The summary statistics of $[Y_L, Y_U]$ for the latter two DGPs are presented in

Table 2.

Table 2. Summary Statistics of DGP2 and DGP3

Brackets	Variable	# of Values	$[\theta_l, \theta_u]$	$[\sigma_l, \sigma_u]$	ρ	Δ
16	$[Y_L, Y_U]$	13290	$[4.4409, 4.9059]$	$[1.10, 0.861]$	0.495	0.4650
2	$[Y_L, Y_U]$	13290	$[3.5283, 7.2534]$	$[1.830, 1.440]$	1.0	3.7251

The length of the identified interval Δ in the 16 bracket case is eight times smaller than that of the 2-bracket case. Moreover, the magnitude of Δ in the 16 bracket experiment is almost half of σ_l and σ_u. So, θ_l and θ_u in the 16 bracket case are close enough for us to expect b_n to play a role at least in small samples. In contrast, in the two bracket case, Δ is large almost twice of $\max\{\sigma_l, \sigma_u\}$.

To implement CI_{FP} and CI_{S}, we need to choose b_n. We used $b_n = s.d. \left(\hat{\Delta}\right) c/\ln(n)$ with $c \in \{0, 3.5, 4\}$. When $c = 0$, $b_n = 0$ which does not satisfy our conditions on b_n in Theorem 1. We chose this b_n to illustrate two points. First, when the parameter θ_0 is point identified or when Δ is small, it's possible that $\hat{\theta}_l$ is larger than $\hat{\theta}_u$ in which case, the effect of using the shrinkage estimator with $b_n = 0$ is to replace negative $\hat{\Delta}$'s with zero; Second, when Δ is large enough, the shrinkage estimator with $b_n = 0$ is the same as the original estimator and in this case, we'll observe the performance of CI_{FP} and CI_{S} using the original estimator $\hat{\Delta}$. When $c = 3.5, 4$, b_n satisfies the conditions of Theorem 1, CI_{FP} and CI_{S} are uniformly asymptotically valid and non-conservative in all cases.

Throughout the simulation, we used $\alpha = 0.05$ and 2000 replications. We compare the finite sample performance of CI_{FP}, CI_{S}, and CI_{AG} via their minimum coverage rates referred to as finite sample confidence sizes, see AG (2007). Given that their asymptotic confidence sizes are achieved at either θ_l $(h_l = 0)$ or θ_u $(h_u = 0)$, we report the respective coverage rates of CI_{FP}, CI_{S}, and CI_{AG} for $\theta = \theta_l, \theta_u$.

Point-Identified Case

We first present results for $Y_{Li} = Y_{Ui}$ for $i = 1, .., n$. In this case, $\hat{\theta}_l = \hat{\theta}_u$, so $\hat{\Delta} = 0$ and all three CIs are the same given by:

$$CI_n = \left[\hat{\theta}_l - \frac{1.96\hat{\sigma}_l}{\sqrt{n}}, \hat{\theta}_l + \frac{1.96\hat{\sigma}_l}{\sqrt{n}}\right].$$

This is also the CI of IM and Horowitz and Manski (2000).

Table 3. Summary Statistics for CI_n

n	CR(θ_0)	Width
500	0.9485	0.1720
1000	0.9525	0.1219
2000	0.950	0.0861
8000	0.9520	0.0431

Its coverage rates denoted by CR(θ_0) and width over 2000 simulations are reported in Table 3. As expected, the coverage rate is very close to the nominal level (0.95) for all sample sizes considered.

In the second experiment, $\{Y_{Li}\}_{i=1}^n \neq \{Y_{Ui}\}_{i=1}^n$, even though $E[Y_{Li}] = E[Y_{Ui}]$. In this case, $\hat{\Delta}$ may not be exactly zero. In fact, it is possible that $\hat{\Delta}$ is negative. Since we drew random samples $\{Y_{Li}\}$ and $\{Y_{Ui}\}$ independently, we would expect this to happen at about 50% of the simulations. In Table 4, we presented the proportion of simulations with $\hat{\Delta} < b_n$ denoted by $P(\Delta^*)$. This is the proportion of simulations in which the shrinkage estimator Δ^* plays a role. When $c = 0$, $P(\Delta^*)$ shows the proportion of simulations with negative $\hat{\Delta}$. It is about 0.5 for all sample sizes. In addition, we reported the coverage rates and width of each CI based on each value of b_n together with the average of $\sqrt{c_{1-\alpha}}$ denoted as Avg($\sqrt{c_{1-\alpha}}$)[5].

Several conclusions emerge from Table 4: First, the confidence sizes of all three

[5]For CS_n^S, we provide $\left(\sqrt{c_{l,1-\alpha}}, \sqrt{c_{u,1-\alpha}}\right)$ which correspond $(c_{l,1-\alpha}, c_{u,1-\alpha})$ in the original Stoye's notation.

Table 4. Summary Statistics when $\rho = 0$

n		c	$P(\Delta^*)$	Avg($\sqrt{c_{1-\alpha}}$)	CR(θ_0)	Width
500	CI_{S}	0	0.497	$(1.8487, 1.8268)$	0.9495	0.1619
		$(3.5, 4)$	1	$(1.9553, 1.9558)$	0.9495	0.1722
	CI_{FP}	0	0.497	1.9087	0.9480	0.1701
		$(3.5, 4)$	1	2.0569	0.9480	0.1833
	CI_{AG}			2.0569	0.9480	0.1833
1000	CI_{S}	0	0.4945	$(1.8476, 1.8318)$	0.9425	0.1146
		$3.5, 4$	1	$(1.9546, 1.9555)$	0.9435	0.1218
	CI_{FP}	0	0.4945	1.9110	0.9430	0.1206
		$(3.5, 4)$	1	2.0569	0.9445	0.1298
	CI_{AG}			2.0569	0.9445	0.1298
2000	CI_{S}	0	0.496	$(1.8459, 1.8323)$	0.9455	0.0806
		$(3.5, 4)$	1	$(1.9551, 1.9547)$	0.9455	0.0857
	CI_{FP}	0	0.496	1.9101	0.9425	0.0849
		$(3.5, 4)$	1	2.0569	0.9425	0.0915
	CI_{AG}			2.0569	0.9425	0.0915
8000	CI_{S}	0	0.499	$(1.844, 1.833)$	0.9470	0.0404
		$(3.5, 4)$	1	$(1.9547, 1.9549)$	0.9470	0.0430
	CI_{FP}	0	0.499	1.9087	0.9480	0.0425
		$(3.5, 4)$	1	2.0568	0.9480	0.0458
	CI_{AG}			2.0568	0.9480	0.0458

CIs are almost the same for all sample sizes and are close to the nominal level, ranging from 0.9421 to 0.9495; Second, the coverage rates of each of CI_{FP} and CI_{S} are almost the same across the three values of c. The one with $c = 0$ shows slightly narrower CI than $c = 3.5, 4$; Third, CI_{FP} with $c = 3.5, 4$ is the same as CI_{AG}, as $P(\Delta^*) = 1$ in both cases; Fourth, the critical values in this case are no longer 1.96 as in the case $\{Y_{Li}\}_{i=1}^{n} = \{Y_{Ui}\}_{i=1}^{n}$, as $\rho = 0$ in this case.

Interval-Identified Case

Sixteen Brackets: Small Δ The coverage rates for θ_l and θ_u along with some summary statistics are presented in Table 5.

In sharp contrast to the point identified case, the confidence sizes of CI_{FP} and CI_{S} in this case differ significantly for $c = 0$ and $c = 3.5, 4$. Note that when $c = 0$, $P(\Delta^*) = 0$,

Table 5. Summary Statistics for 16 Brackets

n		c	$P(\Delta^*)$	Avg($\sqrt{c_{1-\alpha}}$)	Width	CR(θ_l)	CR(θ_u)
500	CI_{S}	0	0	(1.6449, 1.6449)	0.6082	**0.9235**	0.9360
		(3.5, 4)	1	(1.9024, 2.0263)	0.6353	**0.9550**	0.9725
	CI_{FP}	0	0	1.6449	0.6082	**0.9235**	0.9360
		(3.5, 4)	1	1.9759	0.6371	**0.9595**	0.9655
	CI_{AG}			1.9759	0.6371	**0.9595**	0.9655
1000	CI_{S}	0	0	(1.6449, 1.6449)	0.5653	**0.9230**	0.9340
		3.5, 4	1	(1.9020, 2.0260)	0.5845	**0.9535**	0.9715
	CI_{FP}	0	0	1.6449	0.5653	**0.9230**	0.9340
		(3.5, 4)	1	1.9760	0.5857	**0.9570**	0.9630
	CI_{AG}			1.9760	0.5857	**0.9570**	0.9630
2000	CI_{S}	0	0	(1.6449, 1.6449)	0.5367	**0.9335**	0.9370
		3.5	0.4655	(1.7641, 1.8228)	0.5429	**0.9515**	0.9625
		4	1	(1.9015, 2.0263)	0.5503	**0.9570**	0.9685
	CI_{FP}	0	0	1.6449	0.5367	**0.9335**	0.9370
		3.5	0.4655	1.7990	0.5433	**0.9570**	0.9580
		4	1	1.9761	0.5512	0.9640	**0.9630**
	CI_{AG}			1.9761	0.5512	0.9640	**0.9630**
8000	CI_{S}	(0, 3.5, 4)	0	(1.6449, 1.6449)	0.5013	0.9450	**0.9435**
	CI_{FP}	(0, 3.5, 4)	0	1.6449	0.5013	0.9450	**0.9435**
	CI_{AG}			1.9761	0.5086	0.9720	**0.9705**

so the shrinkage estimator didn't play any role in CI_{FP} and CI_{S}. Comparing the confidence sizes of CI_{FP} and CI_{S} for $c = 0$ and $c = 3.5$, we see clearly the role played by the shrinkage estimator Δ^*. When $c = 0$, $P(\Delta^*) = 0$ and both CI_{FP} and CI_{S} under cover except when $n = 8000$, but when $c = 3.5$, $P(\Delta^*) = 1$ for $n = 500, 1000$ and $P(\Delta^*) = 0.4655$ for $n = 2000$, the confidence sizes of both CI_{FP} and CI_{S} are closer to 0.95. When $c = 4$, $P(\Delta^*) = 1$ for $n = 500, 1000, 2000$ and the confidence size of CI_{FP} is the same as that of CI_{AG}. When $n = 8000$, $P(\Delta^*) = 0$ for all c and the confidence size of both CI_{FP} and CI_{S} is 0.9435 as opposed to 0.9705 for CI_{AG}, confirming the non-conservative nature of CI_{FP} and CI_{S}. In general the width of CI_{FP} is slightly larger than that of CI_{S}.

It is very interesting to compare the confidence sizes of CI_{FP} for $c = 0$ across n. For all n, CI_{FP} for $c = 0$ uses the one-sided critical value $\Phi^{-1}(1-\alpha)$. But when $n = 500, 1000, 2000$, $\sqrt{n}\Delta$ is not large enough for the asymptotics to take effect leading to smaller confidence size. In contrast, when $n = 8000$, $\sqrt{n}\Delta$ is large enough leading to the confidence size of 0.9435, the same as the confidence size for $c = 3.5, 4$. These results demonstrate clearly the role of c or b_n when $\sqrt{n}\Delta$ is not large enough (see $n = 500$, e.g.): increase the critical values so as to correct the confidence size. When $\sqrt{n}\Delta$ is large enough, c or b_n is no longer effective and the asymptotics kick in.

Two Brackets: Large Δ In this case, $\sqrt{n}\Delta$ is large enough for all sample sizes considered and b_n does not play any role, i.e., $P(\Delta^*) = 0$ for all c and all sample sizes.

Table 6. Summary Statistics for Two Brackets

n		Avg($\sqrt{c_{1-\alpha}}$)	Width	CR(θ_l)	CR(θ_u)
500	CI_{S}	(1.6449, 1.6449)	3.9655	**0.9435**	0.9580
	CI_{FP}	1.6449	3.9655	**0.9435**	0.9580
	CI_{AG}	1.960	4.0115	**0.9655**	0.9775
1000	CI_{S}	(1.6449, 1.6449)	3.8949	**0.9455**	0.9495
	CI_{FP}	1.6449	3.8949	**0.9455**	0.9495
	CI_{AG}	1.960	3.8949	**0.9685**	0.9785
2000	CI_{S}	(1.6449, 1.6449)	3.8453	**0.9480**	0.9495
	CI_{FP}	1.6449	3.8453	**0.9480**	0.9495
	CI_{AG}	1.960	3.8453	**0.9680**	0.9745
8000	CI_{S}	(1.6449, 1.6449)	3.8753	**0.9465**	0.9515
	CI_{FP}	1.6449	3.8753	**0.9465**	0.9515
	CI_{AG}	1.960	3.8753	0.9760	**0.9735**

The first observation from Table 6 is that CI_{S} and CI_{FP} are identical with confidence size being very close to the nominal level 0.95 for all sample sizes. However, CI_{AG} is quite different from CI_{S} and CI_{FP}: it overcovers for all sample sizes. Secondly, the critical value for CI_{AG} is $\Phi^{-1}(1-\alpha/2) = 1.96$, while that for CI_{S} and CI_{FP} is $\Phi^{-1}(1-\alpha) = 1.645$. Since the critical value for CI_{AG} does not depend on Δ, the reason that the critical value for CI_{AG} is $\Phi^{-1}(1-\alpha/2)$ is because $\hat{\rho} = 1$. See Figure 2. On the other hand, the reason the critical value for CI_{S} and CI_{FP} is 1.645 is because $\sqrt{n}\Delta$ is large enough for all sample sizes considered.

Conclusion and Current Research

In this chapter, we provided a detailed theoretical and numerical study on CIs for interval identified parameters. By inverting a test for the value of the interval identified parameter, we not only developed a new CI, but also established its relationship with existing CIs, including that of IM, Horowitz and Manski (2000), Stoye (2007), and AG (2007). This approach allows straightforward extensions to interval identified parameters for which the estimators of the interval bounds are not asymptotically normally distributed, provided they do not have discontinuity as a function of model parameters. Moreover, we are able to generalize our new CI for interval identified parameters to parameters defined by general moment equalities/inequalities.

The simulation results presented in this chapter support the theoretical finding of Stoye (2007) and the current chapter: it is essential to use the shrinkage estimator of the length of the identified interval or that of the slackness parameters in the general case of parameters defined by moment equalities/inequalities. The shrinkage estimator essentially distinguishes between binding and non-binding moment inequalities.

The CI or CS developed in this chapter has applicability in a wide range of economic/econometric models with partially identified parameters. Moreover, the idea underlying them can be extended to partially identified models for which at least one of the assumptions in this chapter is violated. For example, the validity of CI_{FP} relies on the assumption that the asymptotic distribution of $\left(\hat{\theta}_l, \hat{\theta}_u\right)$ does not have a discontinuity in the model parameters. This may be violated in some applications. One of the authors is currently working on two such cases.

Park (2007a) investigates inference for the distribution of the treatment effects of a binary treatment. Using the same notation as in Example 2, but define $\theta_0 = F_\Delta(\delta)$,

$\theta_l = \sup_y \max(F_1(y) - F_0(y-\delta), 0)$ and $\theta_u = 1 + \inf_y \min(F_1(y) - F_0(y-\delta), 0)$. Then it is known that $\theta_l \leq \theta_0 \leq \theta_u$. Again, with randomized data, F_1 and F_0 are identified and thus θ_l, θ_u are identified. Estimators of θ_l, θ_u can be constructed by replacing F_1 and F_0 with their consistent estimators such as the empirical distributions in the above expressions. However, the estimators of θ_l, θ_u do not satisfy Assumption IM (i), as their asymptotic distribution exhibits discontinuity depending on the value of $\sup_y(F_1(y) - F_0(y-\delta))$ and $\inf_y(F_1(y) - F_0(y-\delta))$. Park (2007b) is an application of this to the Project STAR. Project STAR, conducted by Tennessee State Department of Education in 1985-1988, is a randomized experiment to investigate the effect of class size reduction (CSR) on students' performances. Although the potential heterogeneity of treatment effects of Project STAR has been well-awared (for example, Ding and Lehrer 2005), the heterogeneity has not been fully investigated empirically.

Another extension of the partial identification is Park (2007c). It studies the 'mixing problem' discussed by Manski (1997, 2003). The 'mixing problem' arrises, for example, when we want to "extrapolate the results from a randomized experiment (Manski 2003)". It is because we do not know the 'treatment shares' i.e. the possibility that people comply the rule and do not. When we do not know the 'treatment shares', the probability of a certain range of outcome, say $y \in B$, to occur is bounded in $[\max\{F_1(y \in B) + F_0(y \in B) - 1, 0\},$ $\min\{F_1(y \in B) + F_0(y \in B), 1\}]$, hence the boundary problem at 0 or 1 exist here, too. Park (2007c) studies on the statistical inference of this problem.

Appendix A. Technical Proofs

Proof of Theorem 1

Let

$$\bar{c}_{1-\alpha}(\rho) = \sup_{0 \leq h_l \leq \frac{\sqrt{n}\Delta}{\sigma_l}} cv_{1-\alpha}\left(h_l, \frac{\sqrt{n}\Delta - \sigma_l h_l}{\sigma_u}, \rho\right) \tag{II.15}$$

and

$$\overline{CI}_{\mathrm{FP}} = \left\{\theta : T_n(\theta) \leq \bar{c}_{1-\alpha}(\rho)\right\}.$$

Similar to the proof of Theorem 2 in AG (2007), it is straightforward to show that under Assumption IM (i) and (ii), Assumption A0 and Assumption B0 in AG (2007) are satisfied. As a result, a similar argument to AG (2005b, 2007) yields:

$$\lim_{n\to\infty} \inf_{\theta\in\Theta} \inf_{P:\theta_0(P)=\theta} P\left(\theta_0 \in \overline{CI}_{\mathrm{FP}}\right) = 1-\alpha.$$

Define

$$\begin{aligned} W(h_l) &\equiv (Z_{l,\rho} - h_l)_+^2 + (Z_{u,\rho} + h_u)_-^2 \\ &= (Z_{l,\rho} - h_l)_+^2 + \left(Z_{u,\rho} + \frac{\sqrt{n}\Delta}{\sigma_u} - \frac{\sigma_l}{\sigma_u} h_l\right)_-^2. \end{aligned}$$

Since $W(h_l)$ is convex on $\left[0, \frac{\sqrt{n}\Delta}{\sigma_l}\right]$ a.s., we obtain,

$$\begin{aligned} \sup_{h_l \in \left[0, \frac{\sqrt{n}\Delta}{\sigma_l}\right]} W(h_l) &= \max\left\{W(0), W\left(\frac{\sqrt{n}\Delta}{\sigma_l}\right)\right\} \\ &= \max\left\{(Z_{l,\rho})_+^2 + \left(Z_{u,\rho} + \frac{\sqrt{n}\Delta}{\sigma_u}\right)_-^2, \left(Z_{l,\rho} - \frac{\sqrt{n}\Delta}{\sigma_l}\right)_+^2 + (Z_{u,\rho})_-^2\right\}, \end{aligned}$$

i.e.,

$$\bar{c}_{1-\alpha}(\rho) = \max\left\{cv_{1-\alpha}\left(0, \frac{\sqrt{n}\Delta}{\sigma_u}, \rho\right), cv_{1-\alpha}\left(\frac{\sqrt{n}\Delta}{\sigma_l}, 0, \rho\right)\right\}.$$

We now show that the result holds when $\overline{c}_{1-\alpha}(\rho)$ is replaced with $c^*_{1-\alpha}(\hat{\rho})$. Since $\hat{\sigma}_l$, $\hat{\sigma}_u$, and $\hat{\rho}$ are uniformly consistent estimators of σ_l, σ_u, and ρ respectively, the result holds with

$$\widetilde{c}_{1-\alpha}(\hat{\rho}) = \max\left\{cv_{1-\alpha}\left(0, \frac{\sqrt{n}\Delta}{\hat{\sigma}_u}, \hat{\rho}\right), cv_{1-\alpha}\left(\frac{\sqrt{n}\Delta}{\hat{\sigma}_l}, 0, \hat{\rho}\right)\right\}.$$

Finally we need to justify the use of Δ^*. We follow the same argument as Stoye (2007). Let $c_n = \left(n^{-1/2}b_n\right)^{1/2}$. Then $c_n \to 0$ and $n^{1/2}c_n \to \infty$. We consider two cases: Case I. $\Delta_n \geq c_n$; Case II. $\Delta_n < c_n$.

Case I. $\Delta_n \geq c_n$. In this case, $n^{1/2}\Delta_n \geq n^{1/2}c_n \to \infty$, so either $h_l = \infty$ or $h_u = \infty$ or both. Suppose $h_l = \infty$. Then

$$\begin{aligned}
& \Pr\left[\theta_0 \in CI_{\text{FP}}\right] \\
= \ & \Pr\left[T_n(\theta_0) \leq \max\left\{cv_{1-\alpha}\left(0, \frac{\sqrt{n}\Delta^*}{\hat{\sigma}_u}, \hat{\rho}\right), cv_{1-\alpha}\left(\frac{\sqrt{n}\Delta^*}{\hat{\sigma}_l}, 0, \hat{\rho}\right)\right\}\right] \\
\to \ & \Pr\left[(Z_{l,\rho} - h_l)_+^2 + (Z_{u,\rho} + h_u)_-^2 \leq \max\left\{cv_{1-\alpha}\left(0, \frac{\sqrt{n}\Delta^*}{\hat{\sigma}_u}, \hat{\rho}\right), cv_{1-\alpha}\left(\frac{\sqrt{n}\Delta^*}{\hat{\sigma}_l}, 0, \hat{\rho}\right)\right\}\right] \\
\to \ & \Pr\left[(Z_{u,\rho} + h_u)_-^2 \leq \max\left\{cv_{1-\alpha}\left(0, \frac{\sqrt{n}\Delta^*}{\sigma_u}, \rho\right), cv_{1-\alpha}\left(\frac{\sqrt{n}\Delta^*}{\sigma_l}, 0, \rho\right)\right\}\right] \\
\to \ & \Pr\left[(Z_{u,\rho} + h_u)_-^2 \leq \max\left\{cv_{1-\alpha}(0, \infty, \rho), cv_{1-\alpha}(\infty, 0, \rho)\right\}\right] \\
\geq \ & \Pr\left[(Z_{u,\rho})_-^2 \leq \max\left\{cv_{1-\alpha}(0, \infty, \rho), cv_{1-\alpha}(\infty, 0, \rho)\right\}\right] \\
\geq \ & 1 - \alpha\ ,
\end{aligned}$$

where we have used the result that $\Pr\left[\Delta^* = \hat{\Delta}\right] \to 1$ because of $\Pr\left[\hat{\Delta} > b_n\right] \to 1$. The proof for $h_u = \infty$ is similar. Suppose both $h_l = \infty$ and $h_u = \infty$. Then it is easy to see that $\Pr\left[\theta_0 \in CI_{\text{FP}}\right] \to 1$.

Case II. $\Delta_n < c_n$. In this case, Stoye (2007) shows that $\Delta^* = 0 \leq \Delta$ with probability approaching one. Note that

$$
\begin{aligned}
& \Pr\left[\theta_0 \in CI_{\mathrm{FP}}\right] \\
= \;& \Pr\left[T_n(\theta_0) \leq \max\left\{cv_{1-\alpha}\left(0, \frac{\sqrt{n}\Delta^*}{\hat{\sigma}_u}, \hat{\rho}\right), cv_{1-\alpha}\left(\frac{\sqrt{n}\Delta^*}{\hat{\sigma}_l}, 0, \hat{\rho}\right)\right\}\right] \\
\to \;& \Pr\left[(Z_{l,\rho} - h_l)_+^2 + (Z_{u,\rho} + h_u)_-^2 \leq \max\left\{cv_{1-\alpha}\left(0, \frac{\sqrt{n}\Delta^*}{\sigma_u}, \rho\right), cv_{1-\alpha}\left(\frac{\sqrt{n}\Delta^*}{\sigma_l}, 0, \rho\right)\right\}\right] \\
\geq \;& \Pr\left[(Z_{l,\rho} - h_l)_+^2 + (Z_{u,\rho} + h_u)_-^2 \leq cv_{1-\alpha}(0, 0, \rho)\right] \\
\geq \;& \Pr\left[(Z_{l,\rho})_+^2 + (Z_{u,\rho})_-^2 \leq cv_{1-\alpha}(0, 0, \rho)\right] \\
= \;& 1 - \alpha.
\end{aligned}
$$

The proof is completed by noting that when $\Delta = 0$, $\Pr\left[\theta_0 \in CI_{\mathrm{FP}}\right] \to 1 - \alpha$.

Proof of Theorem 2

We prove the result when $p = 2$. The general case is similar. Similar to the proof of Theorems 2.1, we need to justify the use of $\gamma_1^*(\theta) = \left(\gamma_{1,1}^*(\theta), \gamma_{1,2}^*(\theta)\right)$, where

$$
\gamma_{1,j}^*(\theta) = \begin{cases} \frac{\overline{m}_{n,j}(\theta)}{\hat{\sigma}_j(\theta)} & \text{if } \overline{m}_{n,j}(\theta) > b_n \\ 0 & \text{otherwise} \end{cases}
$$

Let $c_n = \left(n^{-1/2} b_n\right)^{1/2}$. Then $c_n \to 0$ and $n^{1/2} c_n \to \infty$.

Case I. $\gamma_{1,j}(\theta) \geq c_n$**,** $j = 1, 2$. In this case, $n^{1/2}\gamma_{1,j}(\theta) \geq n^{1/2} c_n \to \infty$. Thus,

$$
\begin{aligned}
\Pr\left(\theta_0 \in CS_{\mathrm{MC}}\right) \;&\to\; \Pr\left(\sum_{j=p+1}^{p+v}\left[Z_{h_{2,2},j}\right]^2 \leq cv_{1-\alpha}\left(\infty, \infty, \Omega_n(\theta_0)\right)\right) \\
&=\; 1 - \alpha.
\end{aligned}
$$

Case II. $\gamma_{1,j}(\theta) < c_n$**,** $j = 1, 2$. Similar to Stoye (2007), one can show that

$\gamma^*_{1,j}(\theta) = 0 \leq \gamma_{1,j}$ with probability approaching one. Thus,

$$\begin{aligned}
& \Pr\left(\theta_0 \in CS_{\mathrm{MC}}\right) \\
\to \quad & \Pr\left(\sum_{j=1}^{p}\left[Z_{h_{2,2},j}+h_1\right]_-^2+\sum_{j=p+1}^{p+v}\left[Z_{h_{2,2},j}\right]^2 \leq cv_{1-\alpha}\left(0,0,\Omega_n\left(\theta_0\right)\right)\right) \\
\geq \quad & \Pr\left(\sum_{j=1}^{p}\left[Z_{h_{2,2},j}\right]_-^2+\sum_{j=p+1}^{p+v}\left[Z_{h_{2,2},j}\right]^2 \leq cv_{1-\alpha}\left(0,0,\Omega_n\left(\theta_0\right)\right)\right) \\
= \quad & 1-\alpha.
\end{aligned}$$

Case II. Suppose $\gamma_{1,1}(\theta) < c_n$, but $\gamma_{1,2}(\theta) \geq c_n$. The other case is similar. Then $\gamma^*_{1,1}(\theta) = 0 \leq \gamma_{1,1}$ with probability approaching one and $n^{1/2}\gamma_{1,2}(\theta) \geq n^{1/2}c_n \to \infty$. Thus,

$$\begin{aligned}
& \Pr\left(\theta_0 \in CS_{\mathrm{MC}}\right) \\
\to \quad & \Pr\left(\sum_{j=1}^{p}\left[Z_{h_{2,2},j}+h_1\right]_-^2+\sum_{j=p+1}^{p+v}\left[Z_{h_{2,2},j}\right]^2 \leq cv_{1-\alpha}\left(0,\infty,\Omega_n\left(\theta_0\right)\right)\right) \\
\geq \quad & \Pr\left(\left[Z_{h_{2,2},1}\right]_-^2+\sum_{j=p+1}^{p+v}\left[Z_{h_{2,2},j}\right]^2 \leq cv_{1-\alpha}\left(0,\infty,\Omega_n\left(\theta_0\right)\right)\right) \\
= \quad & 1-\alpha.
\end{aligned}$$

The proof is completed by noting that when all the inequalities are binding, $\Pr\left(\theta_0 \in CS_{\mathrm{MC}}\right) \to 1-\alpha$.

Appendix B. An Expression for $J_h(x)$

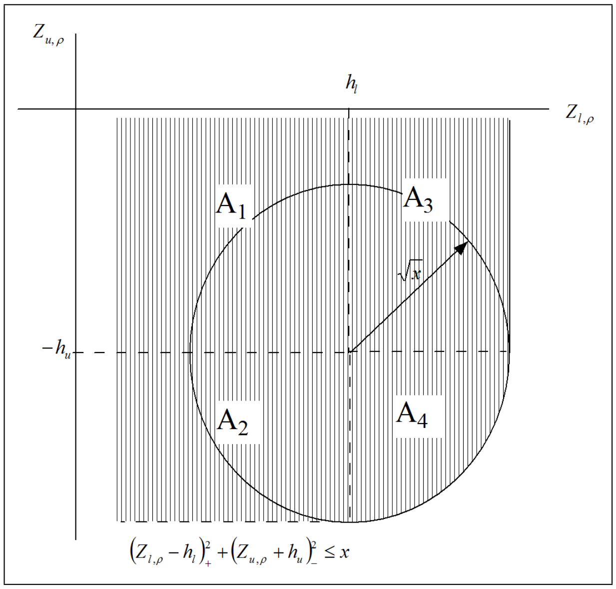
In this section, we derive a closed form expression for $J_h(x)$. This should be useful in constructing CSs in moment inequality models when there are two moment constraints. Let $\phi(z_l, z_u; \rho)$ and $\Phi(z_l, z_u; \rho)$ denote respectively the pdf and cdf of $(Z_{l,\rho}, Z_{u,\rho})$: the standard bivariate normal distribution with correlation coefficient ρ. Define

$$
\begin{aligned}
A_1(x) &= \left\{(z_l, z_u) \in \mathbb{R}^2 : z_l < h_l \text{ and } z_u > -h_u\right\}, \\
A_2(x) &= \left\{(z_l, z_u) \in \mathbb{R}^2 : z_l < h_l \text{ and } -h_u - \sqrt{x} \le z_u \le -h_u\right\}, \\
A_3(x) &= \left\{(z_l, z_u) \in \mathbb{R}^2 : h_l \le z_l \le h_l + \sqrt{x} \text{ and } z_u > -h_u\right\}, \\
A_4(x) &= \Big\{(z_l, z_u) \in \mathbb{R}^2 : h_l \le z_l \le h_l + \sqrt{x}, -h_u - \sqrt{x} \le z_u \le -h_u, \\
&\qquad \text{and } (z_l - h_l)^2 + (z_u + h_u)^2 \le x\Big\}, \\
A(x) &= A_1(x) \cup A_2(x) \cup A_3(x) \cup A_4(x).
\end{aligned}
$$

If $|\rho| < 1$, then

$$
\begin{aligned}
J_h(x) &= J_{(h_l, h_u, \rho)}(x) \\
&= P\left((Z_{l,\rho} - h_l)_+^2 + (Z_{u,\rho} + h_u)_-^2 \le x\right) \\
&= P\left((Z_{l,\rho}, Z_{u,\rho}) \in A_1(x) \cup A_2(x) \cup A_3(x) \cup A_4(x)\right) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I\left\{(z_l, z_u) \in A(x)\right\} \phi(z_l, z_u; \rho)\, dz_l dz_u,
\end{aligned}
$$

where $I(A) = 1$ if A happens; 0 otherwise. Graphically, $A(x)$ is given by the shaded area below.



Hence,

$$
\begin{aligned}
& J_h(x) = \Pr\left[(Z_{l,\rho}-h_l)_+^2 + (Z_{u,\rho}+h_u)_-^2 \le x\right] \\
= \; & \Phi\left(h_l+\sqrt{x}\right) - \Phi\left(h_l, -h_u-\sqrt{x}\right) - \int_{h_l}^{h_l+\sqrt{x}} \int_{-\infty}^{-h_u-\sqrt{x-\left(z_{l,\rho}-h_l\right)^2}} \phi\left(z_l, z_u; \rho\right) dz_u dz_l \\
= \; & \Phi\left(h_l+\sqrt{x}\right) - \int_{-\infty}^{h_l} \phi(z)\, \Phi\left(-\frac{\rho z + h_u + \sqrt{x}}{\sqrt{1-\rho^2}}\right) dz \\
& - \int_{h_l}^{h_l+\sqrt{x}} \phi(z)\, \Phi\left(-\frac{\rho z + h_u + \sqrt{x-(z-h_l)^2}}{\sqrt{1-\rho^2}}\right) dz \\
= \; & \Phi\left(h_l+\sqrt{x}\right) - \int_{-\infty}^{h_l+\sqrt{x}} \phi(z)\, \Phi\left(-\frac{\rho z + h_u + \sqrt{x-(z-h_l)_+^2}}{\sqrt{1-\rho^2}}\right) dz.
\end{aligned}
$$

If $\rho = 1$, then

$$
\left\{(Z_{l,\rho}-h_l)_+^2 + (Z_{u,\rho}+h_u)_-^2 \le x\right\} = \left\{Z : (Z-h_l)_+^2 + (Z+h_u)_-^2 \le x\right\},
$$

where Z is a standard normal random variable. A similar analysis shows that

$$\begin{aligned}
&\left\{Z : (Z-h_l)_+^2 + (Z+h_u)_-^2 \leq x\right\} \\
= &\left\{h_l < Z \leq h_l + \sqrt{x}\right\} \cup \left\{-h_u - \sqrt{x} \leq Z < -h_u\right\} \cup \{-h_u \leq Z \leq h_l\} \\
= &\left\{-h_u - \sqrt{x} < Z \leq h_l + \sqrt{x}\right\}.
\end{aligned}$$

Therefore, we get

$$\begin{aligned}
J_{(h_l,h_u,1)}(x) &= \Pr\left((Z_{l,\rho}-h_l)_+^2 + (Z_{u,\rho}+h_u)_-^2 \leq x\right) \\
&= \Phi\left(h_l + \sqrt{x}\right) - \Phi\left(-h_u - \sqrt{x}\right).
\end{aligned}$$

If $\rho = -1$, then

$$\begin{aligned}
\Pr\left((Z_{l,\rho}-h_l)_+^2 + (Z_{u,\rho}+h_u)_-^2 \leq x\right) &= \Pr\left((Z-h_l)_+^2 + (-Z+h_u)_-^2 \leq x\right) \\
&= \Pr\left((Z-h_l)_+^2 + (Z-h_u)_+^2 \leq x\right).
\end{aligned}$$

Let $\max\{h_l, h_u\} = h_{\max}$ and $\min\{h_l, h_u\} = h_{\min}$.

We can rewrite the event $\left\{(Z-h_l)_+^2 + (Z-h_u)_+^2 \leq x\right\}$ as:

$$\left\{(Z-h_l)_+^2 + (Z-h_u)_+^2 \leq x\right\} = B_1(x) \cup B_2(x) \cup B_3(x) \cup B_4(x),$$

where $B_j(x)$, $j = 1, 2, 3, 4$ correspond to the four possibilities in terms of the signs of $(Z-h_l)$, $(Z-h_u)$. For example,

$$B_1(x) = \left\{Z : Z - h_l > 0,\ Z - h_u > 0, \text{ and } (Z-h_l)_+^2 + (Z-h_u)_+^2 \leq x\right\}.$$

Note that $Z - h_l > 0$ and $Z - h_u > 0$ is equivalent to $Z > h_{\max}$. In this case,

$$\begin{aligned}
& \left\{Z : (Z-h_l)_+^2 + (Z-h_u)_+^2 \leq x\right\} \\
= & \left\{Z : \left(Z - \frac{h_l+h_u}{2}\right)^2 \leq \frac{2x-(h_l-h_u)^2}{4}\right\} \\
= & \left\{Z : Z \leq \frac{h_l+h_u+\sqrt{2x-(h_l-h_u)^2}}{2}\right\} \text{ provided } 2x \geq (h_l-h_u)^2 \\
= & \left\{Z : Z \leq \frac{h_{\max}+h_{\min}+\sqrt{2x-(h_{\max}-h_{\min})^2}}{2}\right\} \text{ provided } 2x \geq (h_{\max}-h_{\min})^2 .
\end{aligned}$$

Also,

$$h_{\max} < \frac{h_{\max}+h_{\min}+\sqrt{2x-(h_{\max}-h_{\min})^2}}{2} \implies (h_{\max}-h_{\min})^2 < x.$$

Therefore, we get

$$\begin{aligned}
B_1(x) &= \left\{Z : h_{\max} < Z \leq \frac{h_{\max}+h_{\min}+\sqrt{2x-(h_{\max}-h_{\min})^2}}{2}\right\} \\
&\quad \text{provided } x > (h_{\max}-h_{\min})^2 , \\
B_1(x) &= \varnothing \text{ if } x \leq (h_{\max}-h_{\min})^2 .
\end{aligned}$$

Similarly, we can show:

$$B_2(x) = \left\{Z : h_{\min} \leq Z < \min\left\{h_{\max}, h_{\min}+\sqrt{x}\right\}\right\}$$

$$B_3(x) = \left\{Z : h_{\min} \leq Z < \min\left\{h_{\max}, h_{\min}+\sqrt{x}\right\}\right\}$$

$$B_4(x) = \{Z : Z \leq h_{\min}\} .$$

Combining them altogether, we get

$$
\begin{aligned}
&\left\{(Z-h_l)_+^2+(Z-h_u)_+^2\le x\right\}\\
=\;&\left(-\infty,\min\left\{h_{\max},h_{\min}+\sqrt{x}\right\}\right)\cup\begin{cases}\varnothing \text{ if } x\le(h_{\max}-h_{\min})^2\\ \left(h_{\max},\frac{h_l+h_u+\sqrt{2x-(h_{\max}-h_{\min})^2}}{2}\right] \text{ otherwise}\end{cases}\\
=\;&\begin{cases}(-\infty,h_{\min}+\sqrt{x}) \text{ if } x\le(h_{\max}-h_{\min})^2\\ \left(-\infty,\frac{h_l+h_u+\sqrt{2x-(h_{\max}-h_{\min})^2}}{2}\right) \text{ otherwise}\end{cases}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\Pr\left((Z_{l,\rho}-h_l)_+^2+(Z_{u,\rho}+h_u)_-^2\le x\right)\\
=\;&\begin{cases}\Phi\left(h_{\min}+\sqrt{x}\right) \text{ if } x\le(h_{\max}-h_{\min})^2\\ \Phi\left(\frac{h_{\max}+h_{\min}+\sqrt{2x-(h_{\max}-h_{\min})^2}}{2}\right) \text{ if } (h_{\max}-h_{\min})^2<x\end{cases}.
\end{aligned}
$$

Appendix C. The Form of the Confidence Set CS_n

In this section, we derive a more explicit form for CS_n:

$$
\begin{aligned}
CS_n &= \{\theta : T_n(\theta) \leq c_{1-\alpha}\} \\
&= \left\{\theta : n\left(\frac{\hat{\theta}_l - \theta}{\hat{\sigma}_l}\right)_+^2 + n\left(\frac{\hat{\theta}_u - \theta}{\hat{\sigma}_u}\right)_-^2 \leq c_{1-\alpha}\right\}.
\end{aligned}
$$

We need to distinguish between two cases. Case I. $\hat{\theta}_l \leq \hat{\theta}_u$ and Case II. $\hat{\theta}_l \geq \hat{\theta}_u$.

For Case I, it is easy to show that

$$
\begin{aligned}
CS_n &= \left\{\theta : \hat{\theta}_l - \sqrt{c_{1-\alpha}}\frac{\hat{\sigma}_l}{\sqrt{n}} \leq \hat{\theta}_l\right\} \cup \left\{\theta : \hat{\theta}_u \leq \theta \leq \hat{\theta}_u + \sqrt{c_{1-\alpha}}\frac{\hat{\sigma}_u}{\sqrt{n}}\right\} \cup \left\{\hat{\theta}_l \leq \theta \leq \hat{\theta}_u\right\} \\
&= \left\{\theta : \hat{\theta}_l - \sqrt{c_{1-\alpha}}\frac{\hat{\sigma}_l}{\sqrt{n}} \leq \hat{\theta}_u + \sqrt{c_{1-\alpha}}\frac{\hat{\sigma}_u}{\sqrt{n}}\right\}.
\end{aligned}
$$

Case II is more complicated. We'll examine it in detail. Note that

$$
CS_n = CS_{n1} \cup CS_{n2} \cup CS_{n3},
$$

where

$$
\begin{aligned}
CS_{n1} &= \left\{\theta : n\left(\frac{\hat{\theta}_l - \theta}{\hat{\sigma}_l}\right)_+^2 + n\left(\frac{\hat{\theta}_u - \theta}{\hat{\sigma}_u}\right)_-^2 \leq c_{1-\alpha},\ \theta \leq \hat{\theta}_u < \hat{\theta}_l\right\}, \\
CS_{n2} &= \left\{\theta : n\left(\frac{\hat{\theta}_l - \theta}{\hat{\sigma}_l}\right)_+^2 + n\left(\frac{\hat{\theta}_u - \theta}{\hat{\sigma}_u}\right)_-^2 \leq c_{1-\alpha},\ \hat{\theta}_u < \hat{\theta}_l \leq \theta\right\}, \\
CS_{n3} &= \left\{\theta : n\left(\frac{\hat{\theta}_l - \theta}{\hat{\sigma}_l}\right)_+^2 + n\left(\frac{\hat{\theta}_u - \theta}{\hat{\sigma}_u}\right)_-^2 \leq c_{1-\alpha},\ \hat{\theta}_u \leq \theta \leq \hat{\theta}_l\right\}.
\end{aligned}
$$

By definition, we obtain

$$\begin{aligned}
CS_{n1} &= \left\{\theta : n\left(\frac{\hat{\theta}_l-\theta}{\hat{\sigma}_l}\right)^2 \le c_{1-\alpha}\right\} \cap \left\{\theta : \theta \le \hat{\theta}_u < \hat{\theta}_l\right\} \\
&= \left\{\theta : \hat{\theta}_l - \sqrt{c_{1-\alpha}}\frac{\hat{\sigma}_l}{\sqrt{n}} \le \theta\right\} \cap \left\{\theta : \theta \le \hat{\theta}_u < \hat{\theta}_l\right\} \\
&= \begin{cases} \left\{\theta : \hat{\theta}_l - \sqrt{c_{1-\alpha}}\frac{\hat{\sigma}_l}{\sqrt{n}} \le \theta \le \hat{\theta}_u\right\} & \text{if } \hat{\theta}_l - \sqrt{c_{1-\alpha}}\frac{\hat{\sigma}_l}{\sqrt{n}} \le \hat{\theta}_u \\ \varnothing & \text{otherwise} \end{cases},
\end{aligned}$$

and

$$\begin{aligned}
CS_{n2} = &\left\{\theta : n\left(\frac{\hat{\theta}_u-\theta}{\hat{\sigma}_u}\right)_-^2 \le c_{1-\alpha}\right\} \cap \left\{\hat{\theta}_u < \hat{\theta}_l \le \theta\right\} \\
= &\left\{\theta : n\left(\frac{\theta-\hat{\theta}_u}{\hat{\sigma}_u}\right)_+^2 \le c_{1-\alpha}\right\} \cap \left\{\hat{\theta}_u < \hat{\theta}_l \le \theta\right\} \\
= &\left\{\theta : \theta \le \hat{\theta}_u + \sqrt{c_{1-\alpha}}\frac{\hat{\sigma}_u}{\sqrt{n}}\right\} \cap \left\{\hat{\theta}_u < \hat{\theta}_l \le \theta\right\} \\
= &\begin{cases} \left\{\theta : \hat{\theta}_l \le \theta \le \hat{\theta}_u + \sqrt{c_{1-\alpha}}\frac{\hat{\sigma}_u}{\sqrt{n}}\right\} & \text{if } \hat{\theta}_l \le \hat{\theta}_u + \sqrt{c_{1-\alpha}}\frac{\hat{\sigma}_u}{\sqrt{n}} \\ \varnothing & \text{otherwise} \end{cases}.
\end{aligned}$$

Now,

$$\begin{aligned}
&CS_{n3} \\
= &\left\{\theta : n\left(\frac{\hat{\theta}_l-\theta}{\hat{\sigma}_l}\right)_+^2 + n\left(\frac{\theta-\hat{\theta}_u}{\hat{\sigma}_u}\right)_+^2 \le c_{1-\alpha}\right\} \cap \left\{\hat{\theta}_u \le \theta \le \hat{\theta}_l\right\} \\
= &\left\{\theta : \left(\hat{\sigma}_u^2+\hat{\sigma}_l^2\right)\theta^2 - 2\left(\hat{\sigma}_u^2\hat{\theta}_l + \hat{\sigma}_l^2\hat{\theta}_u\right)\theta + \hat{\sigma}_u^2\hat{\theta}_l^2 + \hat{\sigma}_l^2\hat{\theta}_u^2 \le \frac{c_{1-\alpha}}{n}\hat{\sigma}_l^2\hat{\sigma}_u^2\right\} \cap \left\{\hat{\theta}_u \le \theta \le \hat{\theta}_l\right\} \\
= &\left\{\theta : \left(\theta - \left(\frac{\hat{\sigma}_u^2\hat{\theta}_l + \hat{\sigma}_l^2\hat{\theta}_u}{\hat{\sigma}_u^2+\hat{\sigma}_l^2}\right)\right)^2 \le \frac{\hat{\sigma}_l^2\hat{\sigma}_u^2}{n\left(\hat{\sigma}_u^2+\hat{\sigma}_l^2\right)}\left[c_{1-\alpha} - \frac{n\left(\hat{\theta}_l-\hat{\theta}_u\right)^2}{\left(\hat{\sigma}_u^2+\hat{\sigma}_l^2\right)}\right]\right\} \cap \left\{\hat{\theta}_u \le \theta \le \hat{\theta}_l\right\}.
\end{aligned}$$

1. If $n\hat{\Delta}^2 > \left(\hat{\sigma}_l^2+\hat{\sigma}_u^2\right)c_{1-\alpha}$, then $CS_{n3} = CS_{n1} = CS_{n2} = \varnothing$. So $CS_n = \varnothing$.

2. If $n\hat{\Delta}^2 \leq \left(\hat{\sigma}_l^2 + \hat{\sigma}_u^2\right) c_{1-\alpha}$, then

$$
\begin{aligned}
& CS_{n3} \\
= & \left\{\theta : \left(\theta - \left(\frac{\hat{\sigma}_u^2\hat{\theta}_l + \hat{\sigma}_l^2\hat{\theta}_u}{\hat{\sigma}_u^2 + \hat{\sigma}_l^2}\right)\right)^2 \leq \frac{\hat{\sigma}_l^2\hat{\sigma}_u^2}{n\left(\hat{\sigma}_u^2 + \hat{\sigma}_l^2\right)}\left[c_{1-\alpha} - \frac{n\left(\hat{\theta}_l - \hat{\theta}_u\right)^2}{\left(\hat{\sigma}_u^2 + \hat{\sigma}_l^2\right)}\right]\right\} \\
& \cap \left\{\theta : \hat{\theta}_u \leq \theta \leq \hat{\theta}_l\right\} \\
= & \{\theta : A \leq \theta \leq B\} \cap \left\{\theta : \hat{\theta}_u \leq \theta \leq \hat{\theta}_l\right\},
\end{aligned}
$$

where

$$
\begin{aligned}
A & \equiv \frac{\hat{\sigma}_u^2\hat{\theta}_l + \hat{\sigma}_l^2\hat{\theta}_u}{\hat{\sigma}_u^2 + \hat{\sigma}_l^2} - \sqrt{\frac{\hat{\sigma}_l^2\hat{\sigma}_u^2}{n\left(\hat{\sigma}_u^2 + \hat{\sigma}_l^2\right)}\left[c_{1-\alpha} - \frac{n\left(\hat{\theta}_l - \hat{\theta}_u\right)^2}{\left(\hat{\sigma}_u^2 + \hat{\sigma}_l^2\right)}\right]}, \\
B & \equiv \frac{\hat{\sigma}_u^2\hat{\theta}_l + \hat{\sigma}_l^2\hat{\theta}_u}{\hat{\sigma}_u^2 + \hat{\sigma}_l^2} + \sqrt{\frac{\hat{\sigma}_l^2\hat{\sigma}_u^2}{n\left(\hat{\sigma}_u^2 + \hat{\sigma}_l^2\right)}\left[c_{1-\alpha} - \frac{n\left(\hat{\theta}_l - \hat{\theta}_u\right)^2}{\left(\hat{\sigma}_u^2 + \hat{\sigma}_l^2\right)}\right]}.
\end{aligned}
$$

Simple algebra shows that $\hat{\theta}_u \leq B$ and $\hat{\theta}_l \geq A$ implying

$$
CS_{n3} = [A, B] \cap \left[\hat{\theta}_u, \hat{\theta}_l\right] = \left[\max\left\{A, \hat{\theta}_u\right\}, \min\left\{B, \hat{\theta}_l\right\}\right].
$$

Now, one can show:

$$
\begin{aligned}
\hat{\theta}_u - A & = \frac{\hat{\sigma}_u^2\hat{\Delta}}{\hat{\sigma}_u^2 + \hat{\sigma}_l^2} + \sqrt{\frac{\hat{\sigma}_l^2\hat{\sigma}_u^2}{n\left(\hat{\sigma}_u^2 + \hat{\sigma}_l^2\right)}\left[c_{1-\alpha} - \frac{n\hat{\Delta}^2}{\left(\hat{\sigma}_u^2 + \hat{\sigma}_l^2\right)}\right]} \\
& = \begin{cases} > 0 \text{ if } c_{1-\alpha} > \frac{n}{\hat{\sigma}_l^2}\hat{\Delta}^2 \\ \leq 0 \text{ if } c_{1-\alpha} \leq \frac{n}{\hat{\sigma}_l^2}\hat{\Delta}^2 \end{cases} \Longrightarrow \begin{cases} \max\left\{A, \hat{\theta}_u\right\} = \hat{\theta}_u \text{ if } \hat{\sigma}_l^2 c_{1-\alpha} > n\hat{\Delta}^2 \\ \max\left\{A, \hat{\theta}_u\right\} = A \text{ if } \hat{\sigma}_l^2 c_{1-\alpha} \leq n\hat{\Delta}^2 \end{cases},
\end{aligned}
$$

and

$$
\begin{aligned}
B-\hat{\theta}_l &= \frac{\hat{\sigma}_l^2\hat{\Delta}}{\hat{\sigma}_u^2+\hat{\sigma}_l^2}+\sqrt{\frac{\hat{\sigma}_l^2\hat{\sigma}_u^2}{n\left(\hat{\sigma}_u^2+\hat{\sigma}_l^2\right)}\left[c_{1-\alpha}-\frac{n\hat{\Delta}^2}{\left(\hat{\sigma}_u^2+\hat{\sigma}_l^2\right)}\right]} \\
&= \begin{cases} >0 \text{ if } c_{1-\alpha}>\frac{n}{\hat{\sigma}_u^2}\hat{\Delta}^2 \\ \leq 0 \text{ if } c_{1-\alpha}\leq\frac{n}{\hat{\sigma}_u^2}\hat{\Delta}^2 \end{cases} \Longrightarrow \begin{cases} \min\left\{B,\hat{\theta}_l\right\}=\hat{\theta}_l \text{ if } \hat{\sigma}_u^2 c_{1-\alpha}>n\hat{\Delta}^2 \\ \min\left\{B,\hat{\theta}_l\right\}=B \text{ if } \hat{\sigma}_u^2 c_{1-\alpha}\leq n\hat{\Delta}^2 \end{cases}.
\end{aligned}
$$

Summarizing, when $n\hat{\Delta}^2\leq\left(\hat{\sigma}_l^2+\hat{\sigma}_u^2\right)c_{1-\alpha}$, we get

$$
\begin{aligned}
CS_n &= \left[\hat{\theta}_l-\sqrt{c_{1-\alpha}}\frac{\hat{\sigma}_l}{\sqrt{n}},\hat{\theta}_u\right]\cup\left[\max\left\{\hat{\theta}_u,A\right\},\min\left\{\hat{\theta}_l,B\right\}\right]\cup\left[\hat{\theta}_l,\hat{\theta}_u+\sqrt{c_{1-\alpha}}\frac{\hat{\sigma}_u}{\sqrt{n}}\right] \\
&= \begin{cases} \left[\hat{\theta}_l-\sqrt{c_{1-\alpha}}\frac{\hat{\sigma}_l}{\sqrt{n}},\hat{\theta}_u+\sqrt{c_{1-\alpha}}\frac{\hat{\sigma}_u}{\sqrt{n}}\right] & \text{if } n\hat{\Delta}^2\leq c_{1-\alpha}\min\left\{\hat{\sigma}_l^2,\hat{\sigma}_u^2\right\} \\ \left[\hat{\theta}_l-\sqrt{c_{1-\alpha}}\frac{\hat{\sigma}_l}{\sqrt{n}},B\right] & \text{if } c_{1-\alpha}\hat{\sigma}_u^2<n\hat{\Delta}^2\leq c_{1-\alpha}\hat{\sigma}_l^2 \\ \left[A,\hat{\theta}_u+\sqrt{c_{1-\alpha}}\frac{\hat{\sigma}_u}{\sqrt{n}}\right] & \text{if } c_{1-\alpha}\hat{\sigma}_l^2<n\hat{\Delta}^2\leq c_{1-\alpha}\hat{\sigma}_u^2 \\ [A,B] & \text{if } c_{1-\alpha}\max\left\{\hat{\sigma}_u^2,\hat{\sigma}_u^2\right\}<n\hat{\Delta}^2\leq c_{1-\alpha}\left(\hat{\sigma}_u^2+\hat{\sigma}_l^2\right) \end{cases}.
\end{aligned}
$$

CHAPTER III

SHARP BOUNDS ON THE DISTRIBUTION OF THE TREATMENT EFFECTS AND THEIR CONFIDENCE INTERVALS

Introduction

Evaluating the effect of a treatment or a program is important in diverse disciplines including social sciences and medical sciences. In medical sciences, randomized clinical trials are often used to evaluate the efficacy of a drug or a procedure in the treatment or prevention of disease. The central problem in the evaluation of a treatment is that any potential outcome that program participants would have received without the treatment is not observed. Because of this missing data problem, most work in the treatment effect literature has focused on the evaluation of various average treatment effects such as the mean of the treatment effects, see the recent book by Lee (2005) for discussion and references. However, empirical evidence strongly suggests that treatment effect heterogeneity prevails in many experiments and various interesting effects of the treatment are missed by the average treatment effects alone, see Djebbari and Smith (2004) who studied heterogeneous program impacts in social experiments such as PROGRESA; Black, Smith, Berger, and Noel (2003) who evaluated the Worker Profiling and Reemployment Services system; and Bitler, Gelbach, and Hoynes (2006) who studied the welfare effact of the change from Aid to Families with Dependent Children (AFDC) to Temporary Assistance for Needy Families (TANF) programs. Other work focusing on treatment effect heterogeneity includes Heckman and Robb (1985), Manski (1990), Imbens and Rubin (1997), Lalonde (1995), Dehejia (1997), Heckman and Smith (1993), Heckman, Smith, and Clements (1997), Lechner (1999), Abadie, Angrist, and Imbens (2002).

When responses to treatment differ among otherwise observationally equivalent subjects, the entire distribution of the treatment effects or other features of the treatment effects than its mean may be of interest. Two approaches have been proposed in the literature to study the distribution of the treatment effects. The first one is the bounding approach originated in Manski (1997a). Assuming monotone treatment response, Manski (1997a) developed sharp bounds on the distribution of the treatment effects. In the second approach, restrictions are imposed on the dependence structure between the potential outcomes such that their joint distribution and the distribution of the treatment effects are identified, see, e.g., Heckman, Smith, and Clements (1997), Biddle, Boden, and Reville (2003), Carneiro, Hansen, and Heckman (2003), Aakvik, Heckman, and Vytlacil (2003), among others.

In this chapter, we take the bounding approach and study the estimation and inference on sharp bounds on the distribution of the treatment effects, which are potentially useful when treatment effects are heterogeneous. Unlike Manski (1997a), we do not assume monotone treatment response. Instead, we assume the marginal distributions of the potential outcomes are identified, but their dependence structure is not. One prominent example of this is provided by ideal randomized experiments. In an ideal randomized experiment, participants of the experiment are randomly assigned to a treatment group and a control group. Because of random assignment, observations on the outcome of participants in the treatment group identify the distribution of the potential outcome with treatment and observations on the outcome of participants in the control group identify the distribution of the potential outcome without treatment, but the two independent random samples do not have any information on the dependence structure between the two potential outcomes.

As a result, neither the joint distribution of the potential outcomes nor the distribution of the treatment effects (defined as the difference between the two potential outcomes) is identified.

Sharp bounds on the joint distribution of the potential outcomes with identified marginals are given by the Frechet-Hoeffding lower and upper bound distributions, see Heckman and Smith (1993), Heckman, Smith, and Clements (1997), and Manski (1997b) for their applications in program evaluation. For randomized experiments, Heckman, Smith, and Clements (1997) proposed nonparametric estimates of the Fréchet-Hoeffding distribution bounds and developed a test for the "common effect" model by testing the lower bound of the variance of the treatment effects. They also suggested an alternative test based on the difference between the quantile functions of the marginal distributions of the potential outcomes referred to as the quantile treatment effects (QTE), see Firpo (2005) or Section 2 for more references.

Sharp bounds on the distribution of the treatment effects—the difference between two potential outcomes with identified marginals—are known in the probability literature. A. N. Kolmogorov posed the question of finding sharp bounds on the distribution of a sum of two random variables with fixed marginal distributions. It was first solved by Makarov (1981) and later by Rüschendorf (1982) and Frank, Nelsen, and Schweizer (1987) using different techniques. Frank, Nelsen, and Schweizer (1987) showed that their proof based on copulas can be extended to more general functions than the sum. Sharp bounds on the respective distributions of a difference, a product, and a quotient of two random variables with fixed marginals can be found in Williamson and Downs (1990). More recently, Denuit, Genest, and Marceau (1999) extended the bounds for the sum to arbitrary dimensions and provided some applications in finance and risk management, see Embrechts, Hoeing, and

Juri (2003) and McNeil, Frey, and Embrechts (2005) for more discussions and additional references.

By making use of the expressions in Williamson and Downs (1990), we propose nonparametric estimators of sharp bounds on the distribution of the treatment effects for randomized experiments and establish their asymptotic properties. It turns out that the asymptotic distributions of these bounds may be discontinuous as functions of the values of the marginal distributions, providing additional examples for which the standard bootstrap with the same sample size may not be asymptotically valid. The failure of the standard bootstrap (bootstrap with the same sample size) in non-regular cases has been pointed out in Andrews (2000), Bickel, Götze, and van Zwet (1997), Beran (1997) and the references therein. Subsampling and fewer-than-n bootstrap have been proposed to rectify the failure of the standard bootstrap, see Andrews (2000), Bickel, Götze, and van Zwet (1997), and Beran (1997) for discussion and references. Subsampling was first proposed by Wu (1990) and extended by Politis and Romano (1994), see Politis, Romano, and Wolf (1999) for more applications of subsampling. Bickel, Götze, and van Zwet (1997) provide numerous examples for which fewer-than-n bootstrap works, while standard bootstrap fails. In this chapter, we apply the fewer-than-n bootstrap (Bickel, Götze, and Zwet (1997) and Bickel and Sakov (2005)) to constructing confidence intervals for these sharp bounds. The finite sample performances of the asymptotics based on the standard normal critical values, the standard bootstrap with the same sample size, and the fewer-than-n bootstrap are compared in a simulation study.

Given sharp bounds on the distribution of the treatment effects, we obtain bounds on the class of D-parameters introduced in Manski (1997a). One example of a D-parameter is any quantile of the treatment effect distribution. In addition, we obtain bounds on the class of D_2-parameters of the treatment effect distribution, see Stoye (2005) or Section 2 for the definition of a D_2-parameter. As pointed out in Stoye (2005), many inequality and risk measures are D_2-parameters. These results shed light on the relation and distinction between QTE and the quantile of the treatment effect distribution.

As an initial investigation of a unified approach to bounding or partially identifying the distribution of the treatment effects, this chapter has focused on randomized experiments. Numerous extensions of the methodologies developed in this chapter are possible and worthwhile. Of immediate concern is the incorporation of covariates into the analysis. We extend sharp bounds in Williamson and Downs (1990) to take into account the presence of covariates under the selection-on-observables assumption commonly used in the treatment effect literature, see, e.g., Rosenbaum and Rubin (1983a, b), Hahn (1998), Heckman, Ichimura, Smith, and Todd (1998), Dehejia and Wahba (1999), among others. In general, taking into account observable covariates tightens the bounds.

The rest of this chapter is organized as follows. In Section 2, we review sharp bounds on the distribution of a difference of two random variables and provide bounds on parameters of the treatment effect distribution that respect either first or second order stochastic dominance.[1] In Section 3, we propose nonparametric estimators of the distribution bounds, establish their asymptotic properties, and describe the fewer-than-n bootstrap procedure we use to construct confidence intervals for the distribution bounds. Results from a detailed simulation study are provided in Section 4. In Section 5, we summarize the asymptotic properties of nonparametric estimators of the distribution of a ratio of two random variables, a measure of the relative treatment effects. Section 6 provides sharp bounds on the treatment effect distribution when covariates are available. Section 7 concludes and

[1] Horowitz and Manski (1995) first used the concept of 'respect stochastic dominance'. Manski (1997a) referred to parameters that respect first order stochastic dominance as D-parameters.

discusses interesting extensions. Proofs are collected in Appendix A. Appendix B presents expressions for the sharp bounds on the distribution of the treatment effects for certain known marginal distributions.

Throughout the chapter, we use $\Longrightarrow$ to denote weak convergence. All the limits are taken as the sample size goes to ∞.

Sharp Bounds on the Distribution of the Treatment Effects and D-Parameters

The notation in this chapter follows the convention in the treatment effect literature. We consider a binary treatment and use Y_1 to denote the potential outcome from receiving treatment and Y_0 the outcome without treatment. Let $F(y_1, y_0)$ denote the joint distribution of Y_1, Y_0 with marginals $F_1(\cdot)$ and $F_0(\cdot)$ respectively.

The characterization theorem of Sklar (1959) implies that there exists a copula[2] $C(u,v)$: $(u,v) \in [0,1]^2$ such that $F(y_1, y_0) = C(F_1(y_1), F_0(y_0))$ for all y_1, y_0. Conversely, for any marginal distributions $F_1(\cdot), F_0(\cdot)$ and any copula function C, the function $C(F_1(y_1), F_0(y_0))$ is a bivariate distribution function with given marginal distributions F_1, F_0. This theorem provides the theoretical foundation for the widespread use of the copula approach in generating multivariate distributions from univariate distributions. For reviews, see Joe (1997) and Nelsen (1999). Since copulas connect multivariate distributions to marginal distributions, the copula approach provides a natural way to study the joint distribution of potential outcomes and the distribution of the treatment effects.

For $(u,v) \in [0,1]^2$, let $C^L(u,v) = \max(u+v-1, 0)$ and $C^U(u,v) = \min(u,v)$ denote the Fréchet-Hoeffding lower and upper bounds for a copula, i.e., $C^L(u,v) \leq C(u,v) \leq$

[2]A copula is a bivariate distribution with uniform marginal distributions on $[0,1]$.

$C^U(u, v)$. Then for any (y_1, y_0), the following inequality holds:

$$C^L(F_1(y_1), F_0(y_0)) \leq F(y_1, y_0) \leq C^U(F_1(y_1), F_0(y_0)). \tag{III.1}$$

The bivariate distribution functions $C^L(F_1(y_1), F_0(y_0))$ and $C^U(F_1(y_1), F_0(y_0))$ are referred to as the Fréchet-Hoeffding lower and upper bounds for bivariate distribution functions with fixed marginal distributions F_1 and F_0. They are distributions of perfectly negatively dependent and perfectly positively dependent random variables respectively, see Nelsen (1999) for more discussions.

Heckman and Smith (1993), Heckman, Smith, and Clements (1997), and Manski (1997b) applied (III.1) in the context of program evaluation. Lee (2002) applied (III.1) to bound correlation coefficients in sample selection models. Fan (2006) developed valid statistical inference procedures for $C^L(F_1(y_1), F_0(y_0))$ and $C^U(F_1(y_1), F_0(y_0))$ based on two independent random samples from $F_1(y_1), F_0(y_0)$ respectively.

Sharp Bounds on the Distribution of the Treatment Effects

Let $\Delta = Y_1 - Y_0$ denote the treatment effect or outcome gain and $F_\Delta(\cdot)$ its distribution function. Given the marginals F_1 and F_0, sharp bounds on the distribution of Δ can be found in Williamson and Downs (1990).

Lemma 1 *Let* $F^L(\delta) = \sup_y \max(F_1(y) - F_0(y-\delta), 0)$ *and* $F^U(\delta) = 1 + \inf_y \min(F_1(y) - F_0(y-\delta), 0)$*. Then* $F^L(\delta) \leq F_\Delta(\delta) \leq F^U(\delta)$*.*

We note the following alternative expressions for $F^L(\delta)$ and $F^U(\delta)$:

$$\begin{aligned} F^L(\delta) &= \max(\sup_y \{F_1(y) - F_0(y-\delta)\}, 0), \qquad \text{(III.2)} \\ F^U(\delta) &= 1 + \min(\inf_y \{F_1(y) - F_0(y-\delta)\}, 0). \end{aligned}$$

At any given value of δ, the bounds $(F^L(\delta), F^U(\delta))$ are informative on the value of $F_\Delta(\delta)$ as long as $[F^L(\delta), F^U(\delta)] \subset [0,1]$. Viewed as an inequality among all possible distribution functions, the sharp bounds $F^L(\delta)$ and $F^U(\delta)$ cannot be improved, because it is easy to show that if either F_1 or F_0 is the degenerate distribution at a finite value, then for all δ, we have $F^L(\delta) = F_\Delta(\delta) = F^U(\delta)$. In fact, given any pair of distribution functions F_1 and F_0, the inequality: $F^L(\delta) \leq F_\Delta(\delta) \leq F^U(\delta)$ cannot be improved, that is, the bounds $F^L(\delta)$ and $F^U(\delta)$ for $F_\Delta(\delta)$ are point-wise best-possible, see Frank, Nelsen, and Schweizer (1987) for a proof of this for a sum of random variables and Williamson and Downs (1990) for a general operation on two random variables.

Lemma 1 implies that the treatment effect distribution F_Δ first order stochastically dominates F^U and is first order stochastically dominated by F^L. Let $\succsim_{FSD}$ denote the first order stochastic dominance relation. Then

$$F^L \succsim_{FSD} F_\Delta \succsim_{FSD} F^U.$$

We note that unlike sharp bounds on the joint distribution of Y_1, Y_0, sharp bounds on the distribution of Δ are not reached at the Fréchet-Hoeffding lower and upper bounds for the distribution of Y_1, Y_0.

Let Y_1', Y_0' be perfectly positively dependent and have the same marginal distributions as Y_1, Y_0 respectively. Let $\Delta' = Y_1' - Y_0'$. Then the distribution of Δ' is given by

$$F_{\Delta'}(\delta) = E1\{Y_1' - Y_0' \leq \delta\} = \int_0^1 1\{F_1^{-1}(u) - F_0^{-1}(u) \leq \delta\} du,$$

where $1\{\cdot\}$ is the indicator function the value of which is 1 if the argument is true, 0 otherwise. Similarly, let Y_1'', Y_0'' be perfectly negatively dependent and have the same marginal distributions as Y_1, Y_0 respectively. Let $\Delta'' = Y_1'' - Y_0''$. Then the distribution of Δ'' is given by

$$F_{\Delta''}(\delta) = E1\{Y_1'' - Y_0'' \leq \delta\} = \int_0^1 1\{F_1^{-1}(u) - F_0^{-1}(1-u) \leq \delta\} du.$$

Interestingly, we show in the next lemma that there exists a second order stochastic dominance relation among the three distributions $F_\Delta, F_{\Delta'}, F_{\Delta''}$. Let $\succsim_{SSD}$ denote the second order stochastic dominance relation.

Lemma 2 *Let $F_\Delta, F_{\Delta'}, F_{\Delta''}$ be defined as above. Then*

$F_{\Delta'} \succsim_{SSD} F_\Delta \succsim_{SSD} F_{\Delta''}$.

Theorem 1 in Stoye (2005) shows that $F_{\Delta'} \succsim_{SSD} F_\Delta$ is equivalent to $E[U(\Delta')] \leq E[U(\Delta)]$ or $E[U(Y_1' - Y_0')] \leq E[U(Y_1 - Y_0)]$ for every convex real-valued function U. Corollary 2.3 in Tchen (1980) implies the conclusion of Lemma 2, see also Cambanis, Simons, and Stout (1976).

Bounds on D-Parameters

The sharp bounds on the treatment effect distribution implies bounds on the class of "D-parameters" introduced in Manski (1997a), see also Manski (2003). One example of "D-parameters" is any quantile of the distribution. Stoye (2005) introduced another class of parameters which measure the dispersion of a distribution, including the variance of the distribution. In this section, we show that sharp bounds can be placed on any dispersion or spread parameter of the treatment effect distribution in this class. For convenience, we restate the definitions of both classes of parameters from Stoye (2005). He refers to the class of "D-parameters" as the class of "D_1-parameters".

Definition 1 *A population statistic θ is a D_1-parameter if it increases weakly with first-*

order stochastic dominance, that is,

$$F \succsim_{FSD} G \text{ implies } \theta(F) \geq \theta(G).$$

Obviously if θ is a D_1-parameter, then Lemma 1 implies:

$$\theta(F^L) \geq \theta(F_\Delta) \geq \theta(F^U).$$

For example, taking θ as a quantile of the treatment effect distribution, we obtain immediately its sharp bounds from Lemma 1. In the following, we will use $G^{-1}(u)$ to denote the generalized inverse of a nondecreasing function G, that is, $G^{-1}(u) = \inf\{x|G(x) \geq u\}$. Then Lemma 1 implies: for $0 \leq q \leq 1$,

$$(F^U)^{-1}(q) \leq F_\Delta^{-1}(q) \leq (F^L)^{-1}(q).$$

For the quantile function of a distribution of a sum of two random variables, expressions for its sharp bounds in terms of quantile functions of the marginal distributions are first established in Makarov (1981). They can also be established via the duality theorem, see Schweizer and Sklar (1983). Using the same tool, one can establish the following expressions for sharp bounds on the quantile function of the distribution of the treatment effects, see Williamson and Downs (1990).

Lemma 3 *For $0 \leq q \leq 1$, $(F^U)^{-1}(q) \leq F_\Delta^{-1}(q) \leq (F^L)^{-1}(q)$, where*

$$\begin{aligned}
(F^L)^{-1}(q) &= \begin{cases} \inf_{u\in[q,1]}[F_1^{-1}(u) - F_0^{-1}(u-q)] & \text{if } q \neq 0 \\ F_1^{-1}(0) - F_0^{-1}(1) & \text{if } q = 0, \end{cases} \\
(F^U)^{-1}(q) &= \begin{cases} \sup_{u\in[0,q]}[F_1^{-1}(u) - F_0^{-1}(1+u-q)] & \text{if } q \neq 1 \\ F_1^{-1}(1) - F_0^{-1}(0) & \text{if } q = 1. \end{cases}
\end{aligned}$$

Like bounds on the distribution of the treatment effects, bounds on the quantile function of Δ are not reached at the Fréchet-Hoeffding bounds for the distribution of (Y_1, Y_0). The following lemma provides simple expressions for the quantile functions of the treatment effects when the potential outcomes are either perfectly positively dependent or perfectly negatively dependent.

Lemma 4 *For $q \in [0,1]$, we have (i) $F_{\Delta'}^{-1}(q) = \left[F_1^{-1}(q) - F_0^{-1}(q)\right]$ if $\left[F_1^{-1}(q) - F_0^{-1}(q)\right]$ is an increasing function of q; (ii) $F_{\Delta''}^{-1}(q) = \left[F_1^{-1}(q) - F_0^{-1}(1-q)\right]$.*

The proof of Lemma 4 follows that of Proposition 3.1 in Embrechts, Hoeing, and Juri (2003). In particular, they showed that for a real valued random variable Z and a function φ increasing and left continuous on the range of Z, it holds that the quantile of $\varphi(Z)$ at quantile level q is given by $\varphi\left(F_Z^{-1}(q)\right)$, where F_Z is the distribution function of Z. For (i), we note that $F_{\Delta'}^{-1}(q)$ equals the quantile of $\left[F_1^{-1}(U) - F_0^{-1}(U)\right]$, where U is a uniform random variable on $[0,1]$. Let $\varphi(U) = F_1^{-1}(U) - F_0^{-1}(U)$. Then $F_{\Delta'}^{-1}(q) = \varphi(q) = F_1^{-1}(q) - F_0^{-1}(q)$ provided that $\varphi(U)$ is an increasing function of U. For (ii), let $\varphi(U) = F_1^{-1}(U) - F_0^{-1}(1-U)$. Then $F_{\Delta''}^{-1}(q)$ equals the quantile of $\varphi(U)$. Since $\varphi(U)$ is always increasing in this case, we get $F_{\Delta''}^{-1}(q) = \varphi(q)$.

Note that the condition in (i) is a necessary condition; without this condition, $\left[F_1^{-1}(q) - F_0^{-1}(q)\right]$ can fail to be a quantile function. Doksum (1974) and Lehmann (1974) used $\left[F_1^{-1}(F_0(y_0)) - y_0\right]$ to measure treatment effects. Recently, $\left[F_1^{-1}(q) - F_0^{-1}(q)\right]$ has been used to study treatment effects heterogeneity and is referred to as the quantile treatment effects (QTE), see e.g., Heckman, Smith, and Clements (1997), Abadie, Angrist, and Imbens (2002), Chernozhukov and Hansen (2005), Firpo (2005), Imbens and Newey (2005), among others, for more discussion and references on the estimation of QTE. Manski (1997a) referred to QTE as ΔD-parameters and the quantile of the treatment effect distribution as $D\Delta$-parameters. Assuming monotone treatment response, Manski (1997a) provided sharp bounds on the quantile of the treatment effect distribution.

It is interesting to note that Lemma 4 (i) shows that QTE equals the quantile function of the treatment effects only when the two potential outcomes are perfectly positively dependent AND QTE is increasing in q. Example 1illustrates a case where QTE is decreasing in q and hence is not the same as the quantile function of the treatment effects even when the potential outcomes are perfectly positively dependent. In contrast to QTE, the quantile of the treatment effect distribution is not identified, but can be bounded, see Lemma 3. At any given quantile level, the lower quantile bound $(F^U)^{-1}(q)$ is the smallest outcome gain (worst case) regardless of the dependence structure between the potential outcomes and should be useful to policy makers. For example, $(F^U)^{-1}(0.5)$ is the minimum gain of at least half of the population.

Definition 2 *A population statistic θ is a D_2-parameter if it increases weakly with second order stochastic dominance, i.e.*

$$F \succsim_{SSD} G \text{ implies } \theta(F) \geq \theta(G).$$

If θ is a D_2-parameter, then Lemma 2 implies

$$\theta(F_{\Delta'}) \leq \theta(F_{\Delta}) \leq \theta(F_{\Delta''}).$$

Stoye (2005) defined the class of D_2-parameters in terms of mean-preserving spread. Since the mean of Δ is identified in our context, the two definitions lead to the same class of D_2-parameters. In contrast to D_1-parameters of the treatment effect distribution, bounds on D_2-parameters of the treatment effect distribution are reached when the potential outcomes are perfectly dependent on each other. One example of D_2-parameters is the variance of the treatment effect Δ. Using results in Cambanis, Simons, and Stout (1976), Heckman, Smith, and Clements (1997) provided bounds on the variance of Δ and proposed a test for the common effect model by testing the value of the lower bound of the variance of Δ. Stoye (2005) presents many other examples of D_2-parameters, including many well-known inequality and risk measures.

Illustrative Example

In this subsection, we provide explicit expressions for sharp bounds on the distribution of the treatment effects and its quantiles when $Y_1 \sim N\left(\mu_1, \sigma_1^2\right)$ and $Y_0 \sim N\left(\mu_0, \sigma_0^2\right)$. In addition, we provide explicit expressions for the distribution of the treatment effects and its quantiles when the potential outcomes are perfectly positively dependent, perfectly negatively dependent, and independent.

Distribution Bounds

Explicit expressions for bounds on the distribution of a sum of two random variables are available for the case where both random variables have the same distribution which includes the uniform, the normal, the Cauchy, and the exponential families, see Alsina (1981), Frank, Nelsen, and Schweizer (1987), and Denuit, Genest, and Marceau (1999). Using the alternative expressions in (III.2), we now derive sharp bounds on the distribution of $\Delta = Y_1 - Y_0$.

First consider the case $\sigma_1 = \sigma_0 = \sigma$. Let $\Phi(\cdot)$ denote the distribution function of the standard normal distribution. Simple algebra shows

$$\begin{aligned} \sup_y \{F_1(y) - F_0(y-\delta)\} &= 2\Phi\left(\frac{\delta - (\mu_1 - \mu_0)}{2\sigma}\right) - 1 \text{ for } \delta > \mu_1 - \mu_0, \\ \inf_y \{F_1(y) - F_0(y-\delta)\} &= 2\Phi\left(\frac{\delta - (\mu_1 - \mu_0)}{2\sigma}\right) - 1 \text{ for } \delta < \mu_1 - \mu_0. \end{aligned}$$

Hence,

$$F^L(\delta) = \begin{cases} 0 & \text{if } \delta < \mu_1 - \mu_0 \\ 2\Phi\left(\frac{\delta - (\mu_1 - \mu_0)}{2\sigma}\right) - 1 & \text{if } \delta \geq \mu_1 - \mu_0 \end{cases} \tag{III.3}$$

$$F^U(\delta) = \begin{cases} 2\Phi\left(\frac{\delta-(\mu_1-\mu_0)}{2\sigma}\right) & \text{if } \delta < \mu_1 - \mu_0 \\ 1 & \text{if } \delta \geq \mu_1 - \mu_0 \end{cases} \tag{III.4}$$

When[3] $\sigma_1 \neq \sigma_0$, we get

$$\begin{aligned} \sup_y \{F_1(y) - F_0(y-\delta)\} &= \Phi\left(\frac{\sigma_1 s - \sigma_0 t}{\sigma_1^2 - \sigma_0^2}\right) + \Phi\left(\frac{\sigma_1 t - \sigma_0 s}{\sigma_1^2 - \sigma_0^2}\right) - 1, \\ \inf_y \{F_1(y) - F_0(y-\delta)\} &= \Phi\left(\frac{\sigma_1 s + \sigma_0 t}{\sigma_1^2 - \sigma_0^2}\right) - \Phi\left(\frac{\sigma_1 t + \sigma_0 s}{\sigma_1^2 - \sigma_0^2}\right) + 1, \end{aligned}$$

where $s = \delta - (\mu_1 - \mu_0)$ and $t = \sqrt{s^2 + (\sigma_1^2 - \sigma_0^2)\ln\left(\frac{\sigma_1^2}{\sigma_0^2}\right)}$. For any δ, one can show that $\sup_y \{F_1(y) - F_0(y-\delta)\} > 0$ and $\inf_y \{F_1(y) - F_0(y-\delta)\} < 0$. As a result,

$$\begin{aligned} F^L(\delta) &= \Phi\left(\frac{\sigma_1 s - \sigma_0 t}{\sigma_1^2 - \sigma_0^2}\right) + \Phi\left(\frac{\sigma_1 t - \sigma_0 s}{\sigma_1^2 - \sigma_0^2}\right) - 1, \\ F^U(\delta) &= \Phi\left(\frac{\sigma_1 s + \sigma_0 t}{\sigma_1^2 - \sigma_0^2}\right) - \Phi\left(\frac{\sigma_1 t + \sigma_0 s}{\sigma_1^2 - \sigma_0^2}\right) + 1. \end{aligned}$$

For comparison purposes, we provide expressions for the distribution F_Δ in three special cases.

Case I. Perfect positive dependence. In this case, Y_0 and Y_1 satisfy $Y_0 = \mu_0 + \frac{\sigma_0}{\sigma_1}Y_1 - \frac{\sigma_0}{\sigma_1}\mu_1$.Therefore,

$$\Delta = \begin{cases} \left(\frac{\sigma_1-\sigma_0}{\sigma_1}\right)Y_1 + \left(\frac{\sigma_0}{\sigma_1}\mu_1 - \mu_0\right), \text{ if } \sigma_1 \neq \sigma_0 \\ \mu_1 - \mu_0, \text{ if } \sigma_1 = \sigma_0. \end{cases}$$

[3]Frank, Nelsen, and Schweizer (1987) provided expressions for the sharp bounds on the distribution of a sum of two normal random variables. We believe there are typos in their expressions, as a direct application of their expressions to our case would lead to different expressions from ours. They are:

$$\begin{aligned} F^L(\delta) &= \Phi\left(\frac{-\sigma_1 s - \sigma_0 t}{\sigma_0^2 - \sigma_1^2}\right) + \Phi\left(\frac{\sigma_0 s - \sigma_1 t}{\sigma_0^2 - \sigma_1^2}\right) - 1, \\ F^U(\delta) &= \Phi\left(\frac{-\sigma_1 s + \sigma_0 t}{\sigma_0^2 - \sigma_1^2}\right) + \Phi\left(\frac{\sigma_0 s + \sigma_1 t}{\sigma_0^2 - \sigma_1^2}\right). \end{aligned}$$

If $\sigma_1 = \sigma_0$, then

$$F_{\Delta}(\delta) = \begin{cases} 0 \text{ and } \delta < \mu_1 - \mu_0 \\ \\ 1 \text{ and } \mu_1 - \mu_0 \leq \delta \end{cases} . \tag{III.5}$$

If $\sigma_1 \neq \sigma_0$, then

$$F_{\Delta}(\delta) = \Phi\left(\frac{\delta - (\mu_1 - \mu_0)}{|\sigma_1 - \sigma_0|}\right).$$

Case II. Perfect negative dependence. In this case, we have $Y_0 = \mu_0 - \frac{\sigma_0}{\sigma_1}Y_1 + \frac{\sigma_0}{\sigma_1}\mu_1$. Hence,

$$\begin{aligned} \Delta &= \frac{\sigma_1 + \sigma_0}{\sigma_1}Y_1 - \left(\frac{\sigma_0}{\sigma_1}\mu_1 + \mu_0\right), \\ F_{\Delta}(\delta) &= \Phi\left(\frac{\delta - (\mu_1 - \mu_0)}{\sigma_1 + \sigma_0}\right). \end{aligned}$$

Case III. Independence. This yields

$$F_{\Delta}(\delta) = \Phi\left(\frac{\delta - (\mu_1 - \mu_0)}{\sqrt{\sigma_1^2 + \sigma_0^2}}\right). \tag{III.6}$$

Figure 3plots the bounds on the distribution F_{Δ} (denoted by F_L and F_U) and the distribution F_{Δ} corresponding to perfect positive dependence, perfect negative dependence, and independence (denoted by F_PPD, F_PND, and F_IND respectively) of potential outcomes for the case $Y_1 \sim N(2,2)$ and $Y_0 \sim N(1,1)$. For notational compactness, we use (F_1, F_0) to signify $Y_1 \sim F_1$ and $Y_0 \sim F_0$ throughout the rest of this chapter.

First we observe from Figure 3 that the bounds in this case are informative at all values of δ and are more informative in the tails of the distribution F_{Δ} than in the middle. In addition, Figure 3 indicates that the distribution of the treatment effects for perfectly positively dependent potential outcomes is most concentrated around its mean 1 implied by

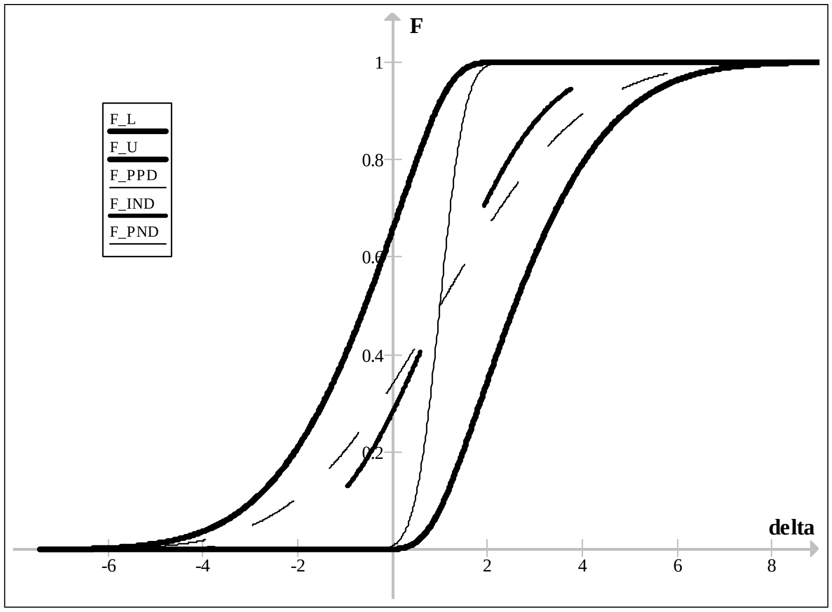


Figure 3. Bounds on the Distribution of the Treatment Effect: $(N(2,2), N(1,1))$

the second order stochastic relation F_PPD $\succsim_{SSD}$F_IND $\succsim_{SSD}$F_PND. In terms of the corresponding quantile functions, this implies that the quantile function corresponding to the perfectly positively dependent potential outcomes is flatter than the quantile functions corresponding to perfectly negatively dependent and independent potential outcomes, see Figure 4.

Quantile Bounds

By inverting (III.3) and (III.4), we obtain the quantile bounds for the case $\sigma_1 = \sigma_0 = \sigma$:

$$\left(F^L\right)^{-1}(q) = \begin{cases} \text{any value in } (-\infty, \mu_1 - \mu_0] \text{ for } q = 0, \\ (\mu_1 - \mu_0) + 2\sigma\Phi^{-1}\left(\frac{1+q}{2}\right) \text{ otherwise;} \end{cases}$$

$$\left(F^U\right)^{-1}(q) = \begin{cases} (\mu_1 - \mu_0) + 2\sigma\Phi^{-1}\left(\frac{q}{2}\right) \text{ for } q \in [0,1), \\ \text{any value in } [\mu_1 - \mu_0, \infty) \text{ for } q = 1. \end{cases}$$

When $\sigma_1 \neq \sigma_0$, there is no closed-form expression for the quantile bounds. But they can be computed numerically by either inverting the distribution bounds or using Lemma 3. We now derive the quantile function for the three special cases.

Case I. Perfect positive dependence. If $\sigma_1 = \sigma_0$, we get

$$F_{\Delta}^{-1}(q) = \begin{cases} \text{any value in } (-\infty, \mu_1 - \mu_0) \text{ for } q = 0, \\ \text{any value in } [\mu_1 - \mu_0, \infty) \text{ for } q = 1, \\ \text{undefined for } q \in (0,1). \end{cases}$$

When $\sigma_1 \neq \sigma_0$, we get

$$F_{\Delta}^{-1}(q) = (\mu_1 - \mu_0) + |\sigma_1 - \sigma_0| \, \Phi^{-1}(q) \text{ for } q \in [0,1].$$

Note that by definition, QTE is given by

$$F_1^{-1}(q) - F_0^{-1}(q) = (\mu_1 - \mu_0) + (\sigma_1 - \sigma_0)\Phi^{-1}(q)$$

which equals $F_{\Delta}^{-1}(q)$ only if $\sigma_1 > \sigma_0$, i.e., only if the condition of Lemma 4 (i) holds. If $\sigma_1 < \sigma_0$, $\left[F_1^{-1}(q) - F_0^{-1}(q)\right]$ is a decreasing function of q and hence can not be a quantile function.

Case II. Perfect negative dependence.

$$F_{\Delta}^{-1}(q) = (\mu_1 - \mu_0) + (\sigma_1 + \sigma_0) \, \Phi^{-1}(q) \text{ for } q \in [0,1].$$

Case III. Independence.

$$F_{\Delta}^{-1}(q) = (\mu_1 - \mu_0) + \sqrt{\sigma_1^2 + \sigma_0^2}\Phi^{-1}(q) \text{ for } q \in [0,1].$$

In Figure 4, we plot the quantile bounds for Δ (FL^{-1} and FU^{-1}) when $Y_1 \sim N(2,2)$ and $Y_0 \sim N(1,1)$ and the quantile functions of Δ when Y_1 and Y_0 are perfectly positively dependent, perfectly negatively dependent, and independent (F_PPD^{-

1}, F_PND^{-1}, and F_IND^{-1} respectively).

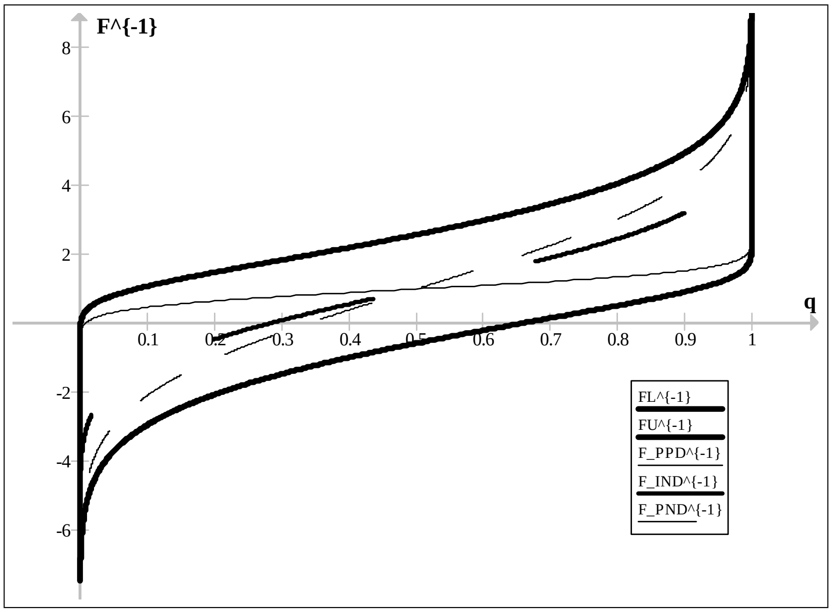


Figure 4. Bounds on the Quantile Function of the Treatment Effect: $(N(2,2), N(1,1))$

Again, Figure 4 reveals the fact that the quantile function of Δ corresponding to the case that Y_1 and Y_0 are perfectly positively dependent is flatter than that corresponding to all the other cases. Keeping in mind that in this case, $\sigma_1 > \sigma_0$, we conclude that the quantile function of Δ in the perfect positive dependence case is the same as QTE. Figure 4 leads to the conclusion that QTE is a conservative measure of the degree of heterogeneity of the treatment effect distribution.

Nonparametric Estimators and Their Asymptotic Properties

Suppose random samples $\{Y_{1i}\}_{i=1}^{n_1} \sim F_1$ and $\{Y_{0i}\}_{i=1}^{n_0} \sim F_0$ are available. Let $\mathcal{Y}_1$ and $\mathcal{Y}_0$ denote respectively the supports[4] of F_1 and F_0. Note that the bounds in Lemma 1

[4]In practice, the supports of F_1 and F_0 may be unknown, but can be estimated by using the corresponding univariate order statistics in the usual way. This won't affect the results to follow. For notational compactness, we assume that they are known.

can be written as

$$F^L(\delta) = \sup_{y\in\mathcal{R}} \left\{F_1(y) - F_0(y-\delta)\right\}, \; F^U(\delta) = 1 + \inf_{y\in\mathcal{R}} \left\{F_1(y) - F_0(y-\delta)\right\}, \tag{III.7}$$

since for any two distributions F_1 and F_0, it is always true that $\sup_{y\in\mathcal{R}} \left\{F_1(y) - F_0(y-\delta)\right\} \geq 0$ and $\inf_{y\in\mathcal{R}} \left\{F_1(y) - F_0(y-\delta)\right\} \leq 0$.

When $\mathcal{Y}_1 = \mathcal{Y}_0 = \mathcal{R}$, (III.7) suggests the following plug-in estimators of $F^L(\delta)$ and $F^U(\delta)$:

$$F_n^L(\delta) = \sup_{y\in\mathcal{R}}\{F_{1n}(y) - F_{0n}(y-\delta)\}, \; F_n^U(\delta) = 1 + \inf_{y\in\mathcal{R}}\{F_{1n}(y) - F_{0n}(y-\delta)\}, \tag{III.8}$$

where $F_{1n}(\cdot)$ and $F_{0n}(\cdot)$ are the empirical distributions defined as

$$F_{kn}(y) = \frac{1}{n_k}\sum_{i=1}^{n_k} 1\left\{Y_{ki} \leq y\right\}, \; k = 1, 0.$$

When either $\mathcal{Y}_1$ or $\mathcal{Y}_0$ is not the whole real line, we derive alternative expressions for $F^L(\delta)$ and $F^U(\delta)$ which turn out to be convenient for both computational purposes and for asymptotic analysis. For illustration, we look at the case: $\mathcal{Y}_1 = \mathcal{Y}_0 = [0,1]$ in detail and provide the results for the general case afterwards.

Suppose $\mathcal{Y}_1 = \mathcal{Y}_0 = [0,1]$. If $1 \geq \delta \geq 0$, then (III.7) implies

$$\begin{aligned}
& F^L(\delta) \\
= \; & \max\left\{ \sup_{y\in[\delta,1]} \left\{F_1(y) - F_0(y-\delta)\right\}, \sup_{y\in(-\infty,\delta)} \left\{F_1(y) - F_0(y-\delta)\right\}, \right. \\
& \left. \sup_{y\in(1,\infty)} \left\{F_1(y) - F_0(y-\delta)\right\} \right\} && \text{(III.9)} \\
= \; & \max\left\{ \sup_{y\in[\delta,1]} \left\{F_1(y) - F_0(y-\delta)\right\}, \sup_{y\in(-\infty,\delta)} F_1(y), \sup_{y\in(1,\infty)} \left\{1 - F_0(y-\delta)\right\} \right\} \\
= \; & \max\left\{ \sup_{y\in[\delta,1]} \left\{F_1(y) - F_0(y-\delta)\right\}, F_1(\delta), 1 - F_0(1-\delta) \right\} \\
= \; & \sup_{y\in[\delta,1]} \left\{F_1(y) - F_0(y-\delta)\right\}, && \text{(III.10)}
\end{aligned}$$

and

$$\begin{aligned}
&F^U(\delta) \\
= \quad & 1+\min\left\{\inf_{y\in[\delta,1]}\left\{F_1(y)-F_0(y-\delta)\right\}, \inf_{y\in(-\infty,\delta)}\left\{F_1(y)-F_0(y-\delta)\right\},\right. \\
& \left.\inf_{y\in(1,\infty)}\left\{F_1(y)-F_0(y-\delta)\right\}\right\} \\
= \quad & 1+\min\left\{\inf_{y\in[\delta,1]}\left\{F_1(y)-F_0(y-\delta)\right\}, \inf_{y\in(-\infty,\delta)} F_1(y), \inf_{y\in(1,\infty)}\left\{1-F_0(y-\delta)\right\}\right\} \\
= \quad & 1+\min\left\{\inf_{y\in[\delta,1]}\left\{F_1(y)-F_0(y-\delta)\right\}, 0\right\};
\end{aligned}$$

If $-1 \leq \delta < 0$, then

$$\begin{aligned}
&F^L(\delta) \\
= \quad & \max\left\{\sup_{y\in[0,1+\delta]}\left\{F_1(y)-F_0(y-\delta)\right\}, \sup_{y\in(-\infty,0)}\left\{F_1(y)-F_0(y-\delta)\right\},\right. \\
& \left.\sup_{y\in(1+\delta,\infty)}\left\{F_1(y)-F_0(y-\delta)\right\}\right\} \qquad \text{(III.11)} \\
= \quad & \max\left\{\sup_{y\in[0,1+\delta]}\left\{F_1(y)-F_0(y-\delta)\right\}, \sup_{y\in(-\infty,0)}\left\{-F_0(y-\delta)\right\}, \sup_{y\in(1+\delta,\infty)}\left\{F_1(y)-1\right\}\right\} \\
= \quad & \max\left\{\sup_{y\in[0,1+\delta]}\left\{F_1(y)-F_0(y-\delta)\right\}, 0\right\}, \qquad \text{(III.12)}
\end{aligned}$$

and

$$\begin{aligned}
&F^U(\delta) \\
= \quad & 1+\min\left\{\inf_{y\in[0,1+\delta]}\left\{F_1(y)-F_0(y-\delta)\right\}, \inf_{y\in(-\infty,0)}\left\{F_1(y)-F_0(y-\delta)\right\},\right. \\
& \left.\inf_{y\in(1+\delta,\infty)}\left\{F_1(y)-F_0(y-\delta)\right\}\right\} \\
= \quad & 1+\min\left\{\inf_{y\in[0,1+\delta]}\left\{F_1(y)-F_0(y-\delta)\right\}, \inf_{y\in(-\infty,0)}\left\{-F_0(y-\delta)\right\}, \inf_{y\in(1+\delta,\infty)}\left\{F_1(y)-1\right\}\right\} \\
= \quad & 1+\inf_{y\in[0,1+\delta]}\left\{F_1(y)-F_0(y-\delta)\right\}.
\end{aligned}$$

Based on (III.10) and (III.12), we propose the following estimator of $F^L(\delta)$:

$$F_n^L(\delta) = \begin{cases} \sup_{y\in[\delta,1]} \{F_{1n}(y) - F_{0n}(y-\delta)\} & \text{if } 1 \geq \delta \geq 0; \\ \max\left\{\sup_{y\in[0,1+\delta]} \{F_{1n}(y) - F_{0n}(y-\delta)\}, 0\right\} & \text{if } -1 \leq \delta < 0. \end{cases}$$

Similarly, we propose the following estimator for $F^U(\delta)$:

$$F_n^U(\delta) = \begin{cases} 1 + \min\left\{\inf_{y\in[\delta,1]} \{F_{1n}(y) - F_{0n}(y-\delta)\}, 0\right\} & \text{if } 1 \geq \delta \geq 0; \\ 1 + \inf_{y\in[0,1+\delta]} \{F_{1n}(y) - F_{0n}(y-\delta)\} & \text{if } -1 \leq \delta < 0. \end{cases}$$

We now summarize the results for general supports $\mathcal{Y}_1$ and $\mathcal{Y}_0$. Suppose $\mathcal{Y}_1 = [a,b]$ and $\mathcal{Y}_0 = [c,d]$ for $a,b,c,d \in \overline{\mathcal{R}} \equiv \mathcal{R} \cup \{-\infty, +\infty\}$, $a < b, c < d$ with $F_1(a) = F_0(c) = 0$ and $F_1(b) = F_0(d) = 1$. It is easy to see that

$$\begin{aligned} F^L(\delta) &= F^U(\delta) = 0, \text{ if } \delta \leq a - d \text{ and} \\ F^L(\delta) &= F^U(\delta) = 1, \text{ if } \delta \geq b - c. \end{aligned}$$

For any $\delta \in [a-d, b-c] \bigcap \mathcal{R}$, let $\mathcal{Y}_\delta = [a,b] \bigcap [c+\delta, d+\delta]$. A similar derivation to the case $\mathcal{Y}_1 = \mathcal{Y}_0 = [0,1]$ leads to

$$F^L(\delta) = \max\left\{\sup_{y\in\mathcal{Y}_\delta} \{F_1(y) - F_0(y-\delta)\}, 0\right\},$$

$$F^U(\delta) = 1 + \min\left\{\inf_{y\in\mathcal{Y}_\delta} \{F_1(y) - F_0(y-\delta)\}, 0\right\},$$

which suggest the following plug-in estimators of $F^L(\delta)$ and $F^U(\delta)$:

$$F_n^L(\delta) = \max\left\{\sup_{y\in\mathcal{Y}_\delta} \{F_{1n}(y) - F_{0n}(y-\delta)\}, 0\right\},$$

$$F_n^U(\delta) = 1 + \min\left\{\inf_{y\in\mathcal{Y}_\delta} \{F_{1n}(y) - F_{0n}(y-\delta)\}, 0\right\}.$$

For any fixed δ, the consistency of $F_n^L(\delta)$ and $F_n^U(\delta)$ is obvious. In the rest of this section, we will establish the asymptotic distributions of $\sqrt{n_1}\left(F_n^L(\delta) - F^L(\delta)\right)$ and

$\sqrt{n_1}\left(F_n^U(\delta) - F^U(\delta)\right)$. By using $F_n^L(\delta)$ and $F_n^U(\delta)$, we can provide bounds on effects of interest other than the average treatment effects including the proportion of people receiving the treatment who benefit from it, see Heckman, Smith, and Clements (1997) for discussion on some of these effects.

Asymptotic Distributions

Define

$$y_{\sup,\delta} = \arg\sup_{y\in\mathcal{Y}_\delta}\left\{F_1(y) - F_0(y-\delta)\right\},\ y_{\inf,\delta} = \arg\inf_{y\in\mathcal{Y}_\delta}\left\{F_1(y) - F_0(y-\delta)\right\},$$

$$M(\delta) = \sup_{y\in\mathcal{Y}_\delta}\left\{F_1(y) - F_0(y-\delta)\right\},\ m(\delta) = \inf_{y\in\mathcal{Y}_\delta}\left\{F_1(y) - F_0(y-\delta)\right\},$$

$$M_n(\delta) = \sup_{y\in\mathcal{Y}_\delta}\left\{F_{1n}(y) - F_{0n}(y-\delta)\right\},\ m_n(\delta) = \inf_{y\in\mathcal{Y}_\delta}\left\{F_{1n}(y) - F_{0n}(y-\delta)\right\}.$$

Then

$$F_n^L(\delta) = \max\left\{M_n(\delta), 0\right\},\ F_n^U(\delta) = 1 + \min\left\{m_n(\delta), 0\right\}.$$

We make the following assumptions.

(A1) (i) The two samples $\{Y_{1i}\}_{i=1}^{n_1}$ and $\{Y_{0i}\}_{i=1}^{n_0}$ are each i.i.d. and are independent of each other; (ii) $n_1/n_0 \to \lambda$ as $n_1 \to \infty$ with $0 < \lambda < \infty$.

(A2) The distribution functions F_1 and F_0 are twice differentiable with bounded density functions f_1 and f_0 on their supports.

(A3) (i) For every $\epsilon > 0$, $\sup_{y\in\mathcal{Y}_\delta:|y-y_{\sup,\delta}|\geq\epsilon}\left\{F_1(y) - F_0(y-\delta)\right\} < \left\{F_1(y_{\sup,\delta}) - F_0(y_{\sup,\delta}-\delta)\right\}$;

(ii) $f_1\left(y_{\sup,\delta}\right) - f_0\left(y_{\sup,\delta}-\delta\right) = 0$ and $f_1'\left(y_{\sup,\delta}\right) - f_0'\left(y_{\sup,\delta}-\delta\right) < 0$.

(A4) (i) For every $\epsilon > 0$, $\inf_{y\in\mathcal{Y}_\delta:|y-y_{\inf,\delta}|\geq\epsilon}\left\{F_1(y) - F_0(y-\delta)\right\} > \left\{F_1(y_{\inf,\delta}) - F_0(y_{\inf,\delta}-\delta)\right\}$;

(ii) $f_1\left(y_{\inf,\delta}\right) - f_0\left(y_{\inf,\delta}-\delta\right) = 0$ and $f_1'\left(y_{\inf,\delta}\right) - f_0'\left(y_{\inf,\delta}-\delta\right) > 0$.

The independence assumption of the two samples in (A1) is satisfied by data from ideal randomized experiments. (A2) imposes smoothness assumptions on the marginal distribution functions. (A3) and (A4) are identifiability assumptions. For a fixed $\delta \in [a-d, b-c] \bigcap \mathcal{R}$, (A3) requires the function $y \longmapsto \{F_1(y) - F_0(y-\delta)\}$ to have a well-separated interior maximum at $y_{\sup,\delta}$ on $\mathcal{Y}_\delta$, while (A4) requires the function $y \longmapsto \{F_1(y) - F_0(y-\delta)\}$ to have a well-separated interior minimum at $y_{\inf,\delta}$ on $\mathcal{Y}_\delta$. If $\mathcal{Y}_\delta$ is compact, then (A3) and (A4) are implied by (A2) and the assumption that the function $y \longmapsto \{F_1(y) - F_0(y-\delta)\}$ have a unique maximum at $y_{\sup,\delta}$ and a unique minimum at $y_{\inf,\delta}$ in the interior of $\mathcal{Y}_\delta$.

We first establish the asymptotic distributions of $M_n(\delta)$ and $m_n(\delta)$.

Proposition 1 *Suppose (A1) and (A2) hold. For a given δ, let*

$$\begin{aligned} \sigma_L^2 &= F_1(y_{\sup,\delta})\left[1 - F_1(y_{\sup,\delta})\right] + \lambda F_0(y_{\sup,\delta} - \delta)\left[1 - F_0(y_{\sup,\delta} - \delta)\right] \text{ and} \\ \sigma_U^2 &= F_1(y_{\inf,\delta})\left[1 - F_1(y_{\inf,\delta})\right] + \lambda F_0(y_{\inf,\delta} - \delta)\left[1 - F_0(y_{\inf,\delta} - \delta)\right]. \end{aligned}$$

Then (i) if (A3) also holds, then$\sqrt{n_1}[M_n(\delta) - M(\delta)] \Longrightarrow N(0, \sigma_L^2)$; *(ii) if (A4) also holds, then* $\sqrt{n_1}[m_n(\delta) - m(\delta)] \Longrightarrow N(0, \sigma_U^2)$.

Following Fan (2006), we obtain immediately Theorem 3by using Proposition 1.

Theorem 3 *(i) Suppose (A1)-(A3) hold. Define* $\infty - \infty = \infty$. *For any* $\delta \in [a-d, b-c] \bigcap \mathcal{R}$, *if* $\min\{a-c, b-d\} < \delta$, *then*$\sqrt{n_1}[F_n^L(\delta) - F^L(\delta)] \Longrightarrow N(0, \sigma_L^2)$; *otherwise*

$$\begin{aligned} \sqrt{n_1}[F_n^L(\delta) - F^L(\delta)] &\Longrightarrow \begin{cases} N\left(0, \sigma_L^2\right) & \text{if } M(\delta) > 0; \\ \max\left\{N(0, \sigma_L^2), 0\right\} & \text{if } M(\delta) = 0; \end{cases} \\ \text{and } \Pr\left(F_n^L(\delta) = 0\right) &\rightarrow 1 \text{ if } M(\delta) < 0. \end{aligned}$$

(ii) Suppose (A1), (A2), and (A4) hold. Define $\infty - \infty = -\infty$. *For any* $\delta \in [a-d, b-c] \bigcap \mathcal{R}$, *if* $\delta < \max\{a-c, b-d\}$, *then*$\sqrt{n_1}[F_n^U(\delta) - F^U(\delta)] \Longrightarrow N(0, \sigma_U^2)$; *otherwise*

$$\begin{aligned} \sqrt{n_1}[F_n^U(\delta) - F^U(\delta)] &\Longrightarrow \begin{cases} N\left(0, \sigma_U^2\right) & \text{if } m(\delta) < 0; \\ \min\left\{N(0, \sigma_U^2), 0\right\} & \text{if } m(\delta) = 0; \end{cases} \\ \text{and } \Pr\left(F_n^U(\delta) = 1\right) &\rightarrow 1 \text{ if } m(\delta) > 0. \end{aligned}$$

Theorem 3 shows that the asymptotic distribution of $F_n^L(\delta)$ ($F_n^U(\delta)$) depends on the value of $M(\delta)$ ($m(\delta)$). For example, if δ is such that $M(\delta) > 0$ ($m(\delta) < 0$), then $F_n^L(\delta)$

$(F_n^U(\delta))$ is asymptotically normally distributed, but if δ is such that $M(\delta) = 0$ $(m(\delta) = 0)$, then the asymptotic distribution of $F_n^L(\delta)$ $(F_n^U(\delta))$ is truncated normal.

Two Examples

We present two examples to illustrate the various possibilities in Theorem 3. For the first example, the asymptotic distribution of $F_n^L(\delta)$ $(F_n^U(\delta))$ is normal for all δ. For the second example, the asymptotic distribution of $F_n^L(\delta)$ $(F_n^U(\delta))$ is normal for some δ and non-normal for some other δ. More examples can be found in Appendix B.

Example 1 (Continued). Let $Y_j \sim N\left(\mu_j, \sigma_j^2\right)$ for $j = 0, 1$ with $\sigma_1^2 \neq \sigma_0^2$. As shown in Section III.2.3, $M(\delta) > 0$ and $m(\delta) < 0$ for all $\delta \in \mathcal{R}$. Moreover,

$$y_{\sup,\delta} = \frac{\sigma_1^2 s + \sigma_1\sigma_0 t}{\sigma_1^2 - \sigma_0^2} + \mu_1 \text{ and } y_{\inf,\delta} = \frac{\sigma_1^2 s - \sigma_1\sigma_0 t}{\sigma_1^2 - \sigma_0^2} + \mu_1$$

are unique interior solutions, where $s = \delta - (\mu_1 - \mu_0)$ and $t = \sqrt{s^2 + 2\left(\sigma_1^2 - \sigma_0^2\right)\ln\frac{\sigma_1}{\sigma_0}}$. Theorem 3 implies that the asymptotic distribution of $F_n^L(\delta)$ $(F_n^U(\delta))$ is normal for all $\delta \in \mathcal{R}$. Inferences can be made using asymptotic distributions or standard bootstrap with the same sample size.

Example 2. Consider the following family of distributions indexed by $a \in (0, 1)$. For brevity, we denote a member of this family by $C\left(a\right)$. If $X \sim C\left(a\right)$, then

$$F(x) = \begin{cases} \frac{1}{a}x^2 \text{ if } x \in [0, a] \\ 1 - \frac{(x-1)^2}{(1-a)} \text{ if } x \in [a, 1] \end{cases} \quad \text{and}$$

$$f(x) = \begin{cases} \frac{2}{a}x \text{ if } x \in [0, a] \\ \frac{2\left(1-x\right)}{(1-a)} \text{ if } x \in [a, 1] \end{cases}.$$

Suppose $Y_1 \sim C\left(\frac{1}{4}\right)$ and $Y_0 \sim C\left(\frac{3}{4}\right)$. The functional form of $F_1\left(y\right) - F_0\left(y - \delta\right)$ differs according to δ. For $y \in \mathcal{Y}_\delta$, using the expressions for $F_1\left(y\right) - F_0\left(y - \delta\right)$ provided in

Appendix B, one can find $y_{\sup,\delta}$ and $M(\delta)$. They are:

$$y_{\sup,\delta} = \begin{cases} \frac{1+\delta}{2} & \text{if } -1+\frac{1}{2}\sqrt{2} < \delta \leq 1 \\ \left\{0, \frac{1+\delta}{2}, 1+\delta\right\} & \text{if } \delta = -1+\frac{1}{2}\sqrt{2} \\ \{0, 1+\delta\} & \text{if } -1 \leq \delta < -1+\frac{1}{2}\sqrt{2} \end{cases};$$

$$M(\delta) = \begin{cases} 4(\delta+1)^2 - 1 & \text{if } -1 \leq \delta \leq -\dfrac{3}{4} \\ -\dfrac{4}{3}\delta^2 & \text{if } -\dfrac{3}{4} \leq \delta \leq -1+\frac{1}{2}\sqrt{2} \\ -\frac{2}{3}(\delta-1)^2 + 1 & \text{if } -1+\frac{1}{2}\sqrt{2} \leq \delta \leq 1 \end{cases}.$$

Figure 5 plots $y_{\sup,\delta}$ and $M(\delta)$ against δ.

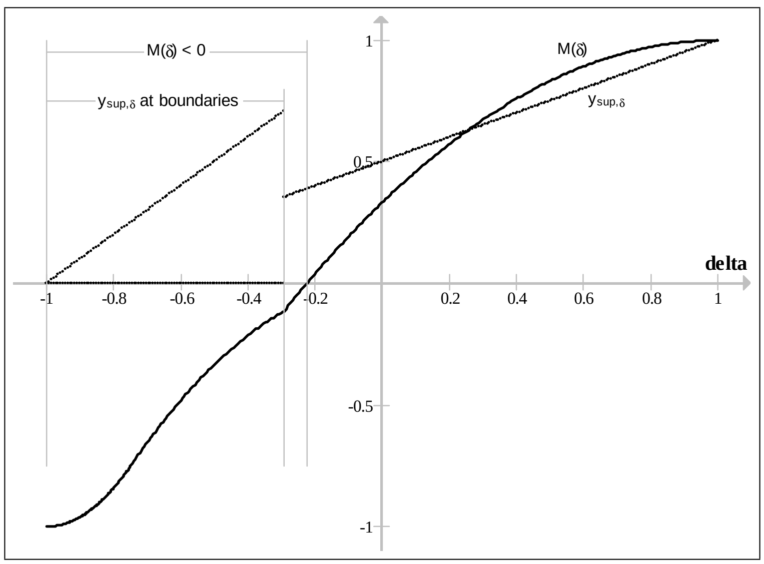


Figure 5. $M(\delta)$ and $y_{\sup,\delta}$: $\left(C\left(\frac{1}{4}\right), C\left(\frac{3}{4}\right)\right)$

Figure 6 plots $F_1(y) - F_0(y-\delta)$ against $y \in [0,1]$ for a few selected values of δ. When $\delta = -\frac{5}{8}$ (Figure 6(a)), the supremum occurs at the boundaries of $\mathcal{Y}_\delta$. When $\delta = -1 + \frac{\sqrt{2}}{2}$ (Figure 6(b)), $\{y_{\sup,\delta}\} = \left\{0, \frac{1+\delta}{2}, 1+\delta\right\}$, i.e., there are three values of $y_{\sup,\delta}$; one interior and two boundary solutions. When $\delta > -1 + \frac{\sqrt{2}}{2}$, $y_{\sup,\delta}$ becomes a unique interior solution. Figure 6(c) plots the case where the interior solution leads to a value 0 for $M(\delta)$ and Figure 6(d) a case where the interior solution corresponds to a positive value for $M(\delta)$.

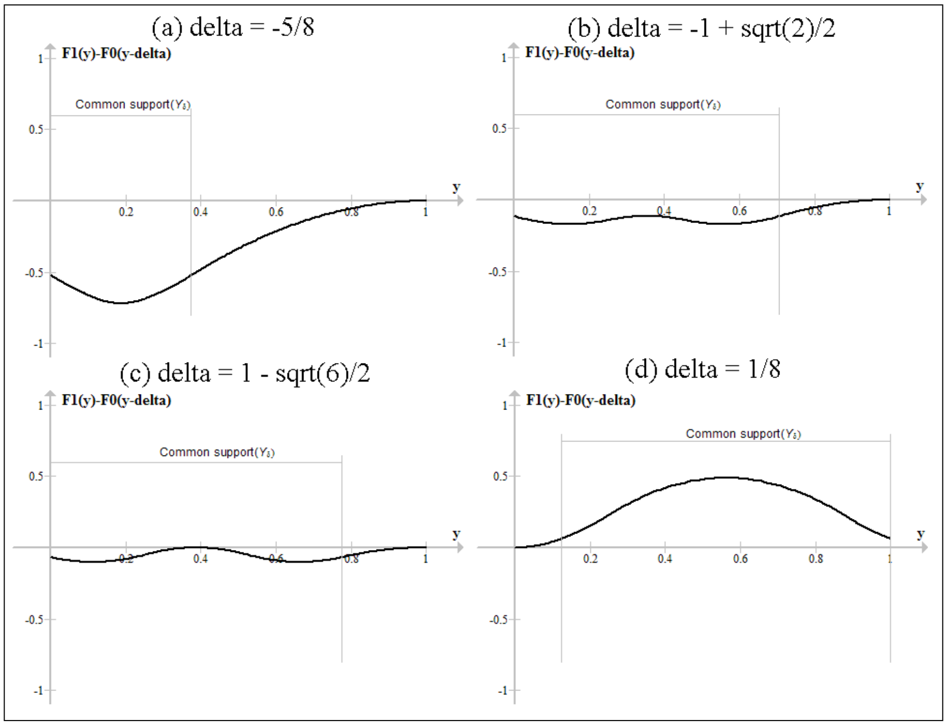


Figure 6. $F_1(y) - F_0(y-\delta)$ for $M(\delta)$

In the simulation study in the next section, we focus on the case of a unique interior solution for $y_{\sup,\delta}$. Depending on the value of δ, $M(\delta)$ can have different signs leading to different asymptotic distributions for $F_n^L(\delta)$. For example, when $\delta = 1 - \frac{\sqrt{6}}{2}$ (Figure 6(c)), $M(\delta) = 0$ and for $\delta > 1 - \frac{\sqrt{6}}{2}$, $M(\delta) > 0$. Since $M(\delta) = 0$ when $\delta = 1 - \frac{\sqrt{6}}{2}$, $y_{\sup,\delta} = 1 - \frac{\sqrt{6}}{4}$ is in the interior, and $f_1'(y_{\sup,\delta}) - f_0'(y_{\sup,\delta} - \delta) = -\frac{16}{3} < 0$, Theorem 3 implies that at $\delta = 1 - \frac{\sqrt{6}}{2}$,

$$\sqrt{n_1}[F_n^L(\delta) - F^L(\delta)] \Longrightarrow \max\left(N(0, \sigma_L^2), 0\right) \text{ where } \sigma_L^2 = \frac{(1+\lambda)}{4}.$$

When $\delta = \frac{1}{8}$ (Figure 6(d)),

$$y_{\sup,\delta} = \frac{9}{16},\ M(\delta) = \frac{47}{96} > 0,\ f_1'(y_{\sup,\delta}) - f_0'(y_{\sup,\delta} - \delta) = -\frac{16}{3} < 0.$$

Theorem 3 implies that when $\delta = \frac{1}{8}$,

$$\sqrt{n_1}[F_n^L(\delta) - F^L(\delta)] \Longrightarrow N(0, \sigma_L^2) \text{ where } \sigma_L^2 = (1+\lambda)\,\frac{7007}{36\,864}.$$

We now illustrate both possibilities for the upper bound $F^U(\delta)$**.** Suppose $Y_1 \sim C\left(\frac{3}{4}\right)$ and $Y_0 \sim C\left(\frac{1}{4}\right)$. Then using the expressions for $F_1(y) - F_0(y-\delta)$ provided in Appendix B, we obtain

$$y_{\inf,\delta} = \begin{cases} \frac{1+\delta}{2} & \text{if } -1 \le \delta \le 1 - \frac{\sqrt{2}}{2} \\ \left\{\delta, \frac{1+\delta}{2}, 1\right\} & \text{if } \delta = 1 - \frac{\sqrt{2}}{2} \\ \{\delta, 1\} & \text{if } 1 - \frac{1}{2}\sqrt{2} \le z \le 1 \end{cases};$$

$$m(\delta) = \begin{cases} \frac{2}{3}(\delta+1)^2 - 1 & \text{if } -1 \le \delta \le 1 - \frac{\sqrt{2}}{2} \\ \dfrac{4}{3}\delta^2 & \text{if } 1 - \frac{\sqrt{2}}{2} \le \delta \le \frac{3}{4} \\ -4(1-\delta)^2 + 1 & \text{if } \frac{3}{4} \le \delta \le 1 \end{cases}.$$

The graphs of $y_{\inf,\delta}$ and $m(\delta)$ are in Figure 7.

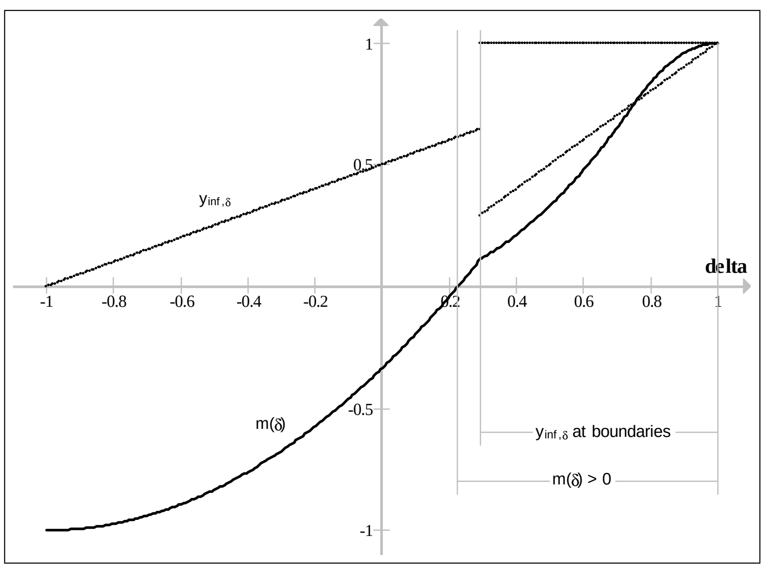


Figure 7. $m(\delta)$ and $y_{\inf,\delta}$: $\left(C\left(\frac{3}{4}\right), C\left(\frac{1}{4}\right)\right)$

Graphs of $F_1(y) - F_0(y-\delta)$ against y for selective δ's are presented in Figure 8.

Figures 8(a) and 8(b) illustrate two cases each having a unique interior minimum, but in Figure 8(a), $m(\delta)$ is negative and in Figure 8(b), $m(\delta)$ is 0. Figure 8(c) illustrates the case with multiple solutions: one interior minimizer and two boundary ones, while Figure 8(d) illustrates the case with two boundary minima.

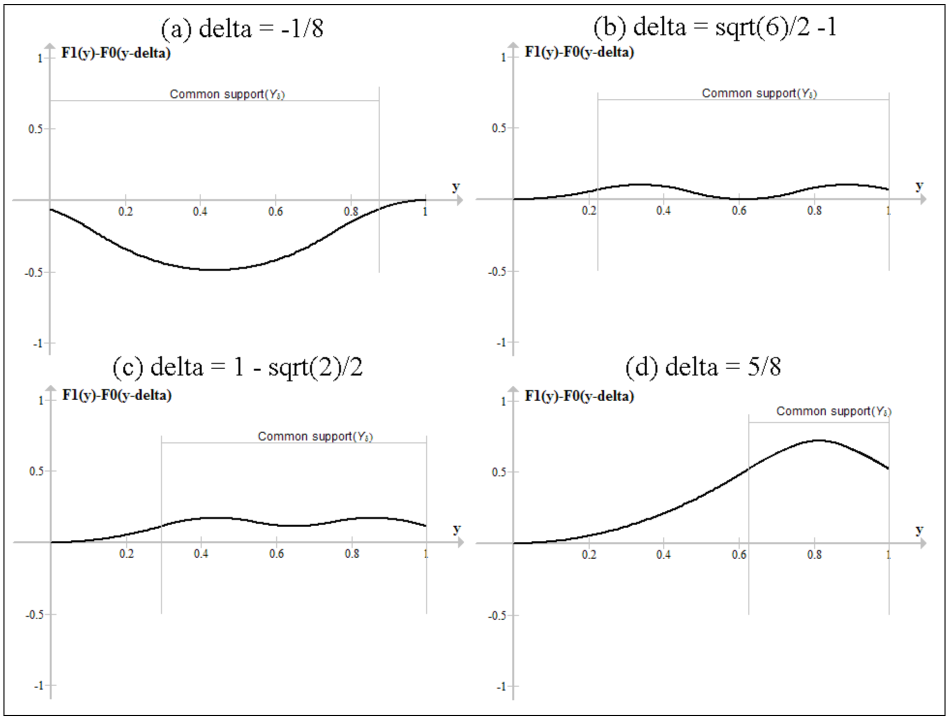


Figure 8. $F_1(y) - F_0(y-\delta)$ for $m(\delta)$

In the simulation study, we considered the case with a unique interior solution corresponding to Figures 8(a) and (b). When $\delta = \dfrac{\sqrt{6}}{2} - 1$, we obtain $y_{\inf,\delta} = \dfrac{\sqrt{6}}{4}$, $m(\delta) = 0$, and $f_1'(y_{\inf,\delta}) - f_0'(y_{\inf,\delta} - \delta) = \dfrac{16}{3} > 0$. By Theorem 3, we get

$$\sqrt{n_1}[F_n^U(\delta) - F^U(\delta)] \Longrightarrow \min\left(N(0, \sigma_U^2), 0\right), \text{ where } \sigma_U^2 = \frac{1+\lambda}{4}.$$

When $\delta = -\dfrac{1}{8}$, we get $y_{\inf,\delta} = \frac{7}{16}$, $m(\delta) = -\frac{47}{96} < 0$, and $f_1'(y_{\inf,\delta}) - f_0'(y_{\inf,\delta} - \delta) = \dfrac{16}{3} > 0$. Hence

$$\sqrt{n_1}[F_n^U(\delta) - F^U(\delta)] \Longrightarrow N(0, \sigma_U^2) \text{ where } \sigma_U^2 = (1+\lambda)\,\frac{7007}{36\,864}.$$

Inference on the Bounds

In the next section, we investigate the performance of the fewer-than-n bootstrap in constructing confidence intervals for $F^L(\delta)$ and $F^U(\delta)$ for δ values corresponding to $y_{\sup,\delta}$ ($y_{\inf,\delta}$) being an interior solution with $M(\delta) > 0$ and $M(\delta) = 0$ ($m(\delta) < 0$ and $m(\delta) = 0$). To implement the fewer-than-n bootstrap, we need to choose the subsample size. We use the procedure suggested in Bickel and Sakov (2005). Let m denote the subsample size and $\hat{m}$ the value of m chosen by the procedure in Bickel and Sakov (2005) (see below for a detailed description of this rule applied to our case). As shown by Bickel and Sakov (2005), $\widehat{m}$ has the desirable property that under general regularity conditions, when the standard bootstrap fails, $\hat{m} \to \infty$ in probability and $\hat{m}/n = o_p(1)$; and when the standard bootstrap works, $\hat{m}/n = O_p(1)$. As a result, there is no loss in efficiency in using the fewer-than-n bootstrap with this adaptive rule of choosing the subsample size. On the other hand, subsampling requires a strictly smaller subsample size.

We now describe this rule for the lower bound $F^L(\delta)$. For notational clarity, we consider the case $n_1 = n_0$. Let $\{Y_{1i}^*\}_{i=1}^m$ be i.i.d. from $F_{1n}(\cdot)$ and $\{Y_{0i}^*\}_{i=1}^m$ i.i.d. from $F_{0n}(\cdot)$ where $m \leq n$. Denote the bootstrap estimators of the sharp bounds by $F_{m,n}^{*L}(\delta)$ and $F_{m,n}^{*U}(\delta)$ and the bootstrap estimators of σ_L^2 and σ_U^2 by $\hat{\sigma}_{m,L}^{2*}$ and $\hat{\sigma}_{m,U}^{2*}$. Let $T_{m,n}^{*LT} = \sqrt{m}\left(F_{m,n}^{*L}(\delta) - F_n^L(\delta)\right)/\hat{\sigma}_{m,L}^*$. To choose m, we follow the steps below.

Step 1. Consider a sequence of m's of the form:

$$m_j = \left[q^j n\right] \text{ for } j = 0, 1, 2, \cdots, \ 0 < q < 1$$

where $[\gamma]$ denotes the largest integer $\leq \gamma$.

Step 2. For each m_j, let $L_{m_j,n}^*$ denote the empirical distribution of values of $T_{m,n}^{*LT}$ over a large number (B) of bootstrap repetitions.

Step 3. Let $\hat{m} = \arg\min_{m_j} \left(\sup_x \left\{ \left| L^*_{m_j,n}(x) - L^*_{m_{j+1},n}(x) \right| \right\} \right)$.

Once $\hat{m}$ is chosen, the confidence intervals can be constructed in the usual way. For example, the $100 * (1-\alpha)\,\%$ two-sided equal-tailed bootstrap confidence interval for $F^L(\delta)$ is

$$\left[F_n^L(\delta) - \frac{1}{n} \frac{c_{\widehat{m},(1-\alpha/2)}}{\hat{\sigma}_L}, F_n^L(\delta) + \frac{1}{n} \frac{c_{\widehat{m},\alpha/2}}{\hat{\sigma}_L} \right],$$

where $c_{m,\beta} = \inf \left\{ x : L^*_{m,n}(x) \geq \beta \right\}$.

Simulation

In this section, we examine the finite sample accuracy of the nonparametric estimators of the treatment effect distribution bounds and investigate the coverage rates of the asymptotic, the standard bootstrap and the fewer-than-n bootstrap confidence intervals for the lower and upper bounds at different values of δ. The data generating processes (DGP) being used in this simulation study are respectively Example 1 and Example 2 introduced in Sections 2 and 3. The detailed simulation design will be described in the subsections below.

Estimates of F^L and F^U

Computation of F_n^L and F_n^U

The quantile functions of F_n^U and F_n^L provide consistent estimators of the lower and upper bounds on the quantile function of F_Δ. For $0 < q < 1$, Lemma 3 (the duality theorem) implies that the quantile bounds $(F_n^U)^{-1}(q)$ and $(F_n^L)^{-1}(q)$ can be computed as follows:

$$\begin{aligned}(F_n^L)^{-1}(q) &= \inf_{u\in[q,1]}\left[F_{1n}^{-1}(u) - F_{0n}^{-1}(u-q)\right],\\ (F_n^U)^{-1}(q) &= \sup_{u\in[0,q]}\left[F_{1n}^{-1}(u) - F_{0n}^{-1}(1+u-q)\right],\end{aligned}$$

where $F_{1n}^{-1}(\cdot)$ and $F_{0n}^{-1}(\cdot)$ represent the quantile functions of $F_{1n}(\cdot)$ and $F_{0n}(\cdot)$ respectively.

To estimate the distribution bounds, we compute the values of $(F_n^L)^{-1}(q)$ and $(F_n^U)^{-1}(q)$ at evenly spaced values of q in $(0,1)$. One choice that leads to easily computed formulas for $(F_n^L)^{-1}(q)$ and $(F_n^U)^{-1}(q)$ is $q = r/n_1$ for $r = 1,\ldots,n_1$, as one can show that

$$(F_n^L)^{-1}(r/n_1) = \min_{l=r,\ldots,(n_1-1)} \min_{s=j,\ldots,k}[Y_{1(l+1)} - Y_{0(s)}], \tag{III.13}$$

where $j = \left[n_0\left(\frac{l-r}{n_1}\right)\right] + 1$ and $k = \left[n_0\left(\frac{l-r+1}{n_1}\right)\right]$, and

$$(F_n^U)^{-1}(r/n_1) = \max_{l=0,\ldots,(r-1)} \max_{s=j',\ldots,k'}[Y_{1(l+1)} - Y_{0(s)}], \tag{III.14}$$

where $j' = \left[n_0\left(\frac{n_1+l-r}{n_1}\right)\right] + 1$ and $k' = \left[n_0\left(\frac{n_1+l-r+1}{n_1}\right)\right]$. In the case where $n_1 = n_0 = n$, (III.13) and (III.14) simplify:

$$\begin{aligned}(F_n^L)^{-1}(r/n) &= \min_{l=r,\ldots,(n-1)}[Y_{1(l+1)} - Y_{0(l-r+1)}],\\ (F_n^U)^{-1}(r/n) &= \max_{l=0,\ldots,(r-1)}[Y_{1(l+1)} - Y_{0(n+l-r+1)}].\end{aligned}$$

The empirical distribution of $(F_n^L)^{-1}(r/n_1)$, $r = 1,...,n_1$ provides an estimate of the lower bound distribution and the empirical distribution of $(F_n^U)^{-1}(r/n_1)$, $r = 1,...,n_1$ provides an estimate of the upper bound distribution.

Simulation Design

The DGPs being used in this experiment are: (i) $F_1 = N(2,1)$ and $F_0 = N(1,1)$; (ii) $F_1 = N(2,2)$ and $F_0 = N(1,1)$; (iii) $F_1 = C(1/4)$ and $F_0 = C(3/4)$; (iv) $F_1 = C(3/4)$

and $F_0 = C(1/4)$. For each set of marginal distributions, random samples of sizes $n_1 = n_0 = n = 1,000$ are drawn and F_n^L and F_n^U are computed. This is repeated for 500 times. Below we present four graphs. In each graph, we plotted F_n^L and F_n^U randomly chosen from the 500 estimates, the averages of 500 F_n^Ls and F_n^Us, and the simulation variances of F_n^L and F_n^U multiplied by n. Each graph consists of 8 curves. The true distribution bounds F^L and F^U are denoted as F^L and F^U, respectively. Their estimates (F_n^L and F_n^U) are Fn^L and Fn^U. The lines denoted by avg(Fn^L) and avg(Fn^U) show the averages of 500 F_n^Ls and F_n^Us. The simulation variances of F_n^L and F_n^U multiplied by n are denoted as n*var(Fn^L) and n*var(Fn^U).

Figures 9(a) and (b) correspond to DGP (i) and (ii), while Figures 10(a) and (b) correspond to DGP (iii) and (iv). In all cases, we observe that Fn^L and avg(Fn^L) are very close to F^L at all points of its support (the same holds true for F^U). In fact, these curves are barely distinguishable from each other. The largest variance in all cases for all values of δ is less than 0.0005.

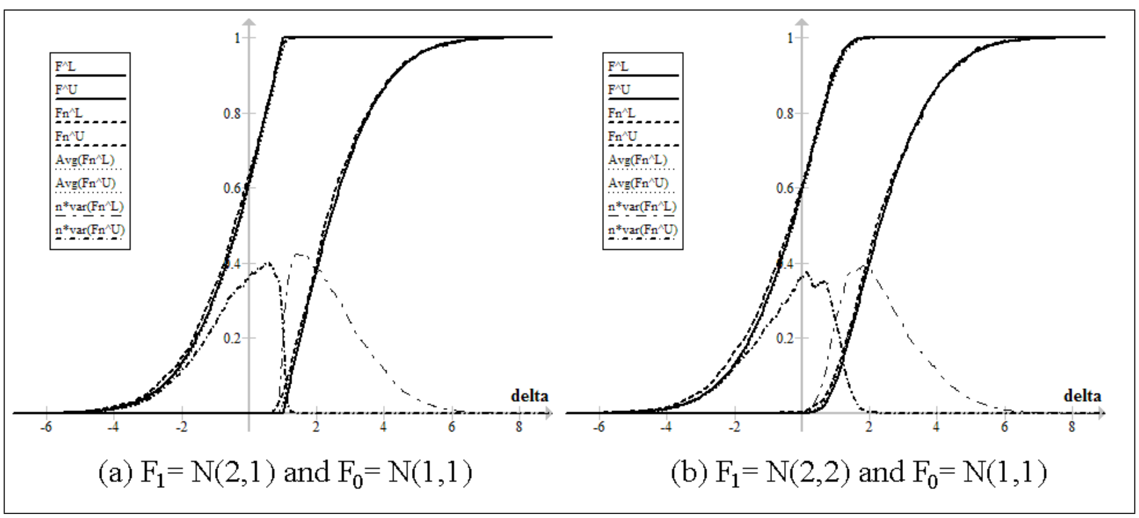


Figure 9. F_n^L and F_n^U for DGP (i) and (ii)

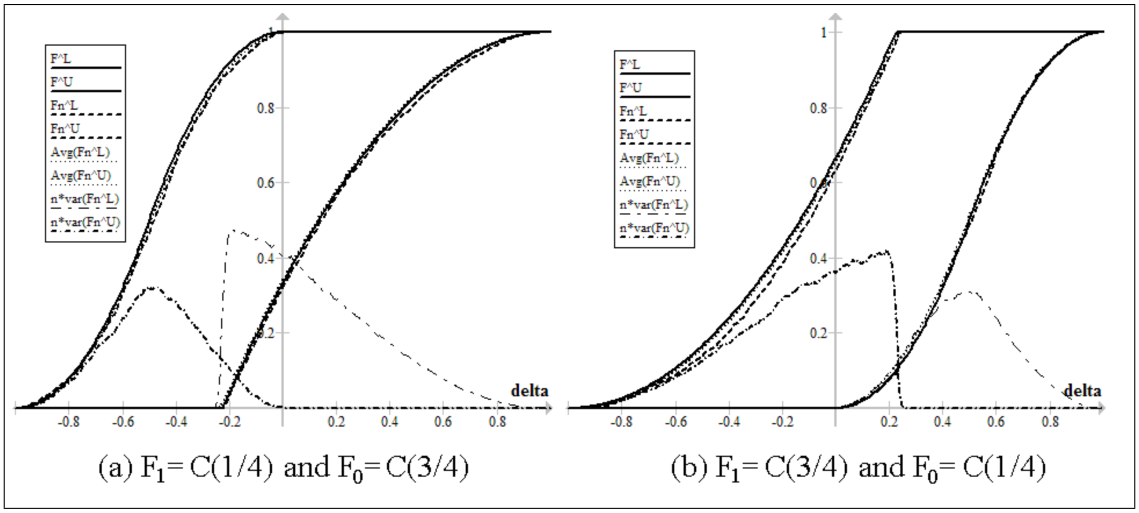


Figure 10. F_n^L and F_n^U for DGP (iii) and (iv)

Coverage Rates

Computation

Construction of the confidence intervals requires estimation of the variances σ_L^2 and σ_U^2 which depend on $y_{\sup.\delta}$ and $y_{\inf.\delta}$. Based on

$$F_n^L(\delta) = \max\{M_n(\delta), 0\} \text{ and } F_n^U(\delta) = 1 + \min\{m_n(\delta), 0\},$$

we now describe a method for computing $M_n(\delta)$, $m_n(\delta)$ and the corresponding $y_{\sup,\delta}$, $y_{\inf,\delta}$.

Suppose we know $\mathcal{Y}_\delta$. To compute $M_n(\delta)$ or $m_n(\delta)$, we just need to consider $Y_{1i} \in \mathcal{Y}_\delta$ and $Y_{0i} \in \mathcal{Y}_\delta - \delta$. If $\mathcal{Y}_\delta$ is unknown, we can estimate it by

$$\mathcal{Y}_{\delta n} = \left[Y_{1(1)}, Y_{1(n_1)}\right] \bigcap \left[Y_{0(1)} + \delta, Y_{0(n_0)} + \delta\right],$$

where $\left\{Y_{1(i)}\right\}_{i=1}^{n_1}$ and $\left\{Y_{0(i)}\right\}_{i=1}^{n_0}$ are the order statistics of $\{Y_{1i}\}_{i=1}^{n_1}$ and $\{Y_{0i}\}_{i=1}^{n_0}$ respectively (in ascending order). In the discussion below, $\mathcal{Y}_\delta$ can be replaced by $\mathcal{Y}_{\delta n}$ if $\mathcal{Y}_\delta$ is unknown.

We define a subset of the order statistics $\{Y_{1i}\}_{i=1}^{n_1}$ denoted as $\left\{Y_{1(i)}\right\}_{i=r_1}^{s_1}$ as follows:

$$r_1 = \arg\min_i \left[\left\{Y_{1(i)}\right\}_{i=1}^{n_1} \bigcap \mathcal{Y}_\delta\right] \text{ and } s_1 = \arg\max_i \left[\left\{Y_{1(i)}\right\}_{i=1}^{n_1} \bigcap \mathcal{Y}_\delta\right].$$

In words, $Y_{1(r_1)}$ is the smallest value of $\left\{Y_{1(i)}\right\}_{i=1}^{n_1} \bigcap \mathcal{Y}_\delta$ and $Y_{1(s_1)}$ is the largest. Then,

$$\begin{aligned} M_n(\delta) &= \max_i \left\{\frac{i}{n_1} - F_{0n}\left(Y_{1(i)} - \delta\right)\right\} \text{ for } i \in \{r_1, r_1+1, \cdots, s_1\} \text{ and} \\ m_n(\delta) &= \min_i \left\{\frac{i}{n_1} - F_{0n}\left(Y_{1(i)} - \delta\right)\right\} \text{ for } i \in \{r_1, r_1+1, \cdots, s_1\}. \end{aligned}$$

To estimate σ_L^2 and σ_U^2, we use the following method. Define two sets I_M and I_m such that

$$\begin{aligned} I_M &= \left\{i : i = \arg\max_i \left\{\frac{i}{n_1} - F_{0n}\left(Y_{1(i)} - \delta\right)\right\}\right\} \text{ and} \\ I_m &= \left\{i : i = \arg\min_i \left\{\frac{i}{n_1} - F_{0n}\left(Y_{1(i)} - \delta\right)\right\}\right\}. \end{aligned}$$

Then the estimators σ_{Ln}^2 and σ_{Un}^2 can be defined as

$$\begin{aligned} \sigma_{Ln}^2 &= \frac{i}{n_1}\left(1 - \frac{i}{n_1}\right) + \lambda F_{0n}\left(Y_{1(i)} - \delta\right)\left(1 - F_{0n}\left(Y_{1(i)} - \delta\right)\right) \text{ and} \\ \sigma_{Un}^2 &= \frac{j}{n_1}\left(1 - \frac{j}{n_1}\right) + \lambda F_{0n}\left(Y_{1(j)} - \delta\right)\left(1 - F_{0n}\left(Y_{1(j)} - \delta\right)\right), \end{aligned}$$

for $i \in I_M$ and $j \in I_m$. Since I_M or I_m may not be singleton, we may have multiple estimates of σ_{Ln}^2 or σ_{Un}^2. In the simulation, we experimented with different ways of selecting σ_{Ln}^2 or σ_{Un}^2 and the results are very similar.

Simulation Results

We looked at pointwise coverage rates of the lower and upper bounds separately at deliberately chosen points. The true marginal distributions and the values of δ used in the simulation are summarized in Table 7.

For Example 1, both Y_1 and Y_0 are normally distributed. As shown in Section III.3.2, $M(\delta) > 0$ and $m(\delta) < 0$ for all three values of δ. Hence the standard bootstrap

Table 7. DGPs Used in the Simulation

	Estimators for	Marginal Distributions		δ		
		F_1	F_0	δ_1	δ_2	δ_3
Example 1	$F^L(\delta)$	$N(2,2)$	$N(1,1)$	1.3	2.6	4.5
	$F^U(\delta)$	$N(2,2)$	$N(1,1)$	-2.4	-0.6	0.7
Example 2	$F^L(\delta)$	$C\left(\frac{1}{4}\right)$	$C\left(\frac{3}{4}\right)$	$\frac{1}{8}$	$1-\frac{\sqrt{6}}{2}$	.
	$F^U(\delta)$	$C\left(\frac{3}{4}\right)$	$C\left(\frac{1}{4}\right)$	$-\frac{1}{8}$	$\frac{\sqrt{6}}{2}-1$	.

works for all δ's. The values of δ are chosen such that $F^L(\delta_1) \approx F^U(\delta_1) \approx 0.15$, $F^L(\delta_2) \approx F^U(\delta_2) \approx 0.5$, and $F^L(\delta_3) \approx F^U(\delta_3) \approx 0.85$ to see the effect of the relative position of δ on the coverage rates. For Example 2, $M(\delta_1) > 0$ and $m(\delta_1) < 0$ while $M(\delta_2) = m(\delta_2) = 0$ for both $F^L(\cdot)$ and $F^U(\cdot)$. Hence the standard bootstrap works for δ_1 but not for δ_2.

For each DGP described in Table 7, we generated random samples of the same size n from F_1 and F_0 respectively. The sample sizes are $n = 1,000, 2,000, 4,000$ and the number of simulations was 1000. To select the number of bootstrap repetitions B, we followed Davidson and Mackinnon (2004; pp163-165) by choosing B such that $\alpha(B+1)$ is an integer. Specifically, we used $B = 999$ for $\alpha = 0.05$. For Example 1, we constructed confidence intervals for $F^L(\delta)$ and $F^U(\delta)$ for each δ by three methods. The first is the confidence interval based on the standard normal distribution. We denote the corresponding results by 'Asymptotics' in Table 8 below. The second method used the standard bootstrap confidence intervals and the results are denoted by 'n-bootstrap' in Table 8. Finally, we used the 'fewer-than-n-bootstrap' confidence intervals. In the 'fewer-than-n-bootstrap', we used $q = 0.95$. Here only one value for q was used, because the 'fewer-than-n bootstrap' was only used for comparison purposes (the standard bootstrap works for this case). For Example 2, we used the standard normal distribution ('Asymptotics' in Table 9), the standard bootstrap ('n-bootstrap' in Table 9), and the 'fewer-than-n-bootstrap' with two values for q: 0.75 and 0.95.

Table 8. Coverage Rates for Example 1

n	Method	$F^L(\delta)$			$F^U(\delta)$		
		δ_1	δ_2	δ_3	δ_1	δ_2	δ_3
$1,000$	Asymptotics	.929	.944	.937	.931	.949	.926
	n-bootstrap	.942	.954	.950	.950	.953	.939
$q = 0.95$	Fewer-than-n bootstrap	.948	.949	.948	.952	.951	.942
$2,000$	Asymptotics	.942	.944	.934	.943	.946	.927
	n-bootstrap	.949	.944	.946	.946	.952	.937
$q = 0.95$	Fewer-than-n bootstrap	.941	.944	.952	.949	.950	.939
$4,000$	Asymptotics	.935	.953	.936	.949	.949	.928
	n-bootstrap	.945	.957	.953	.951	.952	.936
$q = 0.95$	Fewer-than-n bootstrap	.944	.957	.952	.951	.952	.939

Table 9. Coverage Rates for Example 2

n	Method	$F^L(\delta)$		$F^U(\delta)$	
		δ_1	δ_2	δ_1	δ_2
$1,000$	Asymptotics	.933	.935	.947	.935
	n-bootstrap	.941	.961	.951	.958
	Fewer-than-n bootstrap ($q = 0.75$)	.943	.963	.951	.960
	Fewer-than-n bootstrap ($q = 0.95$)	.945	.963	.947	.962
$2,000$	Asymptotics	.952	.955	.940	.940
	n-bootstrap	.951	.970	.947	.959
	Fewer-than-n bootstrap ($q = 0.75$)	.944	.971	.946	.959
	Fewer-than-n bootstrap ($q = 0.95$)	.951	.969	.946	.959
$4,000$	Asymptotics	.948	.944	.952	.946
	n-bootstrap	.947	.963	.946	.963
	Fewer-than-n bootstrap ($q = 0.75$)	.949	.964	.947	.965
	Fewer-than-n bootstrap ($q = 0.95$)	.949	.962	.951	.961

First, we discuss the coverage rates for normal distributions in Table 8. Clearly the coverage rates depend critically on the location of δ. For δ_2, all three methods lead to confidence intervals with very accurate coverage rates for both F^L and F^U. The coverage rates at δ_1 and δ_3 depend on the methods being used. Although in theory all three methods are asymptotically valid, in finite samples, confidence intervals based on normal critical values often substantially under cover the true values at δ_1 and/or δ_3. For example, the coverage rates of confidence intervals based on normal critical values for $F^L(\delta)$ at $\delta = \delta_1$ and δ_3 are respectively .929 and .937 when $n = 1,000$ and .935 and .936 when $n = 4,000$. On the other hand, the standard bootstrap leads to coverage rates of .942 and .950 when

$n = 1,000$ and .945 and .953 when $n = 4,000$, supporting the asymptotic refinement of the standard bootstrap over asymptotic normality in this case. The fewer-than-n bootstrap delivers similar coverage rates to the standard bootstrap.

Table 9 provides the results for Example 2. All three methods: the 'Asymptotics' based on normal critical values, the n-bootstrap and the fewer-than-n bootstrap with different values of q perform similarly at δ_1 except when $n = 1000$, the 'Asymptotics' undercovers for $F^L(\delta_1)$ with coverage rate .933. At δ_2, the n-bootstrap leads to coverage rates higher than .95 for almost all sample sizes, while the fewer-than-n bootstrap produces coverage rates that are slightly better than the n-bootstrap, but not by much. On the other hand, the 'Asymptotics' provides coverage rates that are closer to .95 except when $n = 1000$.

Table 10. MSE and Bias for Example 1

		$F^L(\delta)$			$F^U(\delta)$		
n	Statistics	δ_1	δ_2	δ_3	δ_1	δ_2	δ_3
$1,000$	$\sqrt{MSE}$	.0209	.0194	.0118	.0123	.0198	.0215
	Bias	.0094	.0070	.0046	$-.0054$	$-.0082$	$-.0107$
$2,000$	$\sqrt{MSE}$	.0143	.0135	.0083	.0086	.0138	.0149
	Bias	.0064	.0040	.0033	$-.0030$	$-.0052$	$-.0077$
$4,000$	$\sqrt{MSE}$	.0102	.0094	.0060	.0062	.0097	.0103
	Bias	.0045	.0028	.0022	$-.0022$	$-.0034$	$-.0053$

Table 11. MSE and Bias for Example 2

		$F^L(\delta)$		$F^U(\delta)$	
n	Statistics	δ_1	δ_2	δ_1	δ_2
$1,000$	$\sqrt{MSE}$	.0202	.0216	.0204	.0221
	Bias	.0080	.0147	$-.0087$	$-.0155$
$2,000$	$\sqrt{MSE}$	.0139	.0149	.0144	.0153
	Bias	.0044	.0101	$-.0057$	$-.0104$
$4,000$	$\sqrt{MSE}$	.0098	.0102	.0100	.0103
	Bias	.0033	.0069	$-.0033$	$-.0069$

Tables 10 and 11 present the bias and RMSE of $F_n^L(\delta)$ and $F_n^U(\delta)$ for the values of δ used to evaluate coverage rates. As expected, as the sample size n increases, both the bias and the MSE of the lower and upper bound estimators decrease regardless of the values of δ for both examples. Also for both examples, the lower bound estimator $F_n^L(\delta)$ is biased upward and the upper bound estimator $F_n^U(\delta)$ is biased downward for all sample sizes and for all values of δ considered.

Estimation and Inference on the Distribution of the Relative Treatment Effects

When the potential outcomes are almost surely positive, an alternative measure of the treatment effects is the relative risk defined as the ratio of the two potential outcomes. Let $R = \frac{Y_1}{Y_0}$. A value of R larger than 1 indicates effectiveness of the treatment and a value of R smaller than 1 indicates ineffectiveness of the treatment. Williamson and Downs (1990) showed that the sharp bounds on the distribution of R are:

$$\begin{aligned} F_R^L(\delta) &= \sup_y \max(F_1(y) - F_0(y/\delta), 0) \text{ and} \\ F_R^U(\delta) &= 1 + \inf_y \min(F_1(y) - F_0(y/\delta), 0). \end{aligned}$$

Let $\mathcal{Y}_1 = [a, b]$ and $\mathcal{Y}_0 = [c, d]$ for $a, b, c, d \in \mathcal{R}_+ \bigcup \{0, \infty\}$ denote the supports of Y_1 and Y_0 respectively. Define $\mathcal{Y}_{\delta,R} = [a, b] \bigcap [\delta c, \delta d]$ for $\delta \in \left[\frac{a}{d}, \frac{b}{c}\right] \bigcap \mathcal{R}_+$ with obvious definitions of $\frac{a}{d}$ and $\frac{b}{c}$ when one or more of $a, b, c, d \in \{0, \infty\}$. Then it can be shown that

$$\begin{aligned} F_R^L(\delta) &= \max\left\{ \sup_{y \in \mathcal{Y}_{\delta,R}} [F_1(y) - F_0(y/\delta)], 0 \right\} \equiv \max(M_R(\delta), 0) \text{ and} \\ F_R^U(\delta) &= 1 + \min\left\{ \inf_{y \in \mathcal{Y}_{\delta,R}} [F_1(y) - F_0(y/\delta)], 0 \right\} \equiv 1 + \min(m_R(\delta), 0), \end{aligned}$$

where

$$\begin{aligned} M_R(\delta) &= F_1(y_{\sup,\delta R}) - F_0(y_{\sup,\delta R}/\delta) \text{ and} \\ m_R(\delta) &= F_1(y_{\inf,\delta R}) - F_0(y_{\inf,\delta R}/\delta) \end{aligned}$$

in which

$$\begin{aligned} y_{\sup,\delta R} &= \arg\sup_{y\in\mathcal{Y}_{\delta,R}} \left(F_1(y) - F_0(y/\delta)\right) \text{ and} \\ y_{\inf,\delta R} &= \arg\inf_{y\in\mathcal{Y}_{\delta,R}} \left(F_1(y) - F_0(y/\delta)\right). \end{aligned}$$

Consistent estimators of $F_R^L(\delta)$ and $F_R^U(\delta)$ are:

$$\begin{aligned} F_{nR}^L(\delta) &= \max\left\{\sup_{y\in\mathcal{Y}_{\delta,R}} \left(F_{1n}(y) - F_{0n}(y/\delta)\right), 0\right\} \text{ and} \\ F_{nR}^U(\delta) &= 1 + \min\left\{\inf_{y\in\mathcal{Y}_{\delta,R}} \left(F_{1n}(y) - F_{0n}(y/\delta)\right), 0\right\}. \end{aligned}$$

To provide the asymptotic distributions of $F_{nR}^L(\delta)$ and $F_{nR}^U(\delta)$, we modify (A3) and (A4) to (A3R) and (A4R) below.

(A3R) (i) For every $\epsilon > 0$,

$$\sup_{y\in\mathcal{Y}_{\delta,R}:|y-y_{\sup,\delta R}|\geq\epsilon} \{F_1(y) - F_0(y/\delta)\} < \{F_1(y_{\sup,\delta R}) - F_0(y_{\sup,\delta R}/\delta)\};$$

(ii) $f_1\left(y_{\sup,\delta R}\right) - \frac{1}{\delta} f_0\left(y_{\sup,\delta R}/\delta\right) = 0$ and $f_1'\left(y_{\sup,\delta R}\right) - \frac{1}{\delta^2} f_0'\left(y_{\sup,\delta R}/\delta\right) < 0$.

(A4R) (i) For every $\epsilon > 0$,

$$\inf_{y\in\mathcal{Y}_{\delta,R}:|y-y_{\inf,\delta R}|\geq\epsilon} \{F_1(y) - F_0(y/\delta)\} > \{F_1(y_{\inf,\delta R}) - F_0(y_{\inf,\delta R}/\delta)\};$$

(ii) $f_1\left(y_{\inf,\delta R}\right) - \frac{1}{\delta} f_0\left(y_{\inf,\delta R}/\delta\right) = 0$ and $f_1'\left(y_{\inf,\delta R}\right) - \frac{1}{\delta^2} f_0'\left(y_{\inf,\delta R}/\delta\right) > 0$.

Theorem 4 *(i) Suppose (A1),(A2) and (A3R) hold. Define* $\frac{0}{0} = \frac{\infty}{\infty} = \infty$. *For any* $\delta \in \left[\frac{a}{d}, \frac{b}{c}\right] \bigcap \mathcal{R}_+$, *if* $\min\left\{\frac{a}{c}, \frac{b}{d}\right\} < \delta$, *then* $\sqrt{n_1}[F_{nR}^L(\delta) - F_R^L(\delta)] \Longrightarrow N(0, \sigma_{LR}^2)$; *otherwise*

$$\begin{aligned} \sqrt{n_1}[F_{nR}^L(\delta) - F_R^L(\delta)] &\Longrightarrow \begin{cases} N\left(0, \sigma_{LR}^2\right) & \text{if } M_R(\delta) > 0; \\ \max\left\{N\left(0, \sigma_{LR}^2\right), 0\right\} & \text{if } M_R(\delta) = 0; \end{cases} \\ \text{and } \Pr\left(F_{nR}^L(\delta) = 0\right) &\to 1 \text{ if } M_R(\delta) < 0, \end{aligned}$$

where

$$\sigma_{LR}^2 = F_1(y_{\sup,\delta R})\left[1 - F_1(y_{\sup,\delta R})\right] + \lambda F_0(y_{\sup,\delta R}/\delta)\left[1 - F_0(y_{\sup,\delta R}/\delta)\right].$$

(ii) Suppose (A1), (A2), and (A4R) hold. Define $\frac{0}{0} = \frac{\infty}{\infty} = 0$. *For any* $\delta \in \left[\frac{a}{d}, \frac{b}{c}\right] \bigcap \mathcal{R}_+$, *if* $\max\left\{\frac{a}{c}, \frac{b}{d}\right\} > \delta$, *then* $\sqrt{n_1}[F_{nR}^U(\delta) - F_R^U(\delta)] \Longrightarrow N(0, \sigma_{UR}^2)$; *otherwise*

$$\begin{aligned} \sqrt{n_1}[F_{nR}^U(\delta) - F_R^U(\delta)] &\Longrightarrow \begin{cases} N\left(0, \sigma_{UR}^2\right) & \text{if } m_R(\delta) < 0; \\ \max\left\{N\left(0, \sigma_{UR}^2\right), 0\right\} & \text{if } M_R(\delta) = 0; \end{cases} \\ \text{and } \Pr\left(F_{nR}^U(\delta) = 1\right) &\rightarrow 1 \text{ if } m_R(\delta) > 0, \end{aligned}$$

where

$$\sigma_{UR}^2 = F_1(y_{\inf,\delta R})\left[1 - F_1(y_{\inf,\delta R}/\delta)\right] + \lambda F_0(y_{\inf,\delta R}/\delta)\left[1 - F_0(y_{\inf,\delta R}/\delta)\right].$$

The proof of Theorem 4 is similar to that of Theorem 3 and is thus omitted. Like Theorem 3, Theorem 4 implies that in general, the standard asymptotics and bootstrap may fail to provide valid inference on the sharp bounds $F_R^L(\delta)$ and $F_R^U(\delta)$. Instead the fewer-than-n bootstrap or subsampling should be used to make inferences on these bounds.

Sharp Bounds on the Distribution of Treatment Effects with Covariates

In many applications, observations on a vector of covariates for individuals in the treatment and control groups are available. In this section, we extend our study on sharp bounds to take into account these covariates. For notational compactness, we let $n = n_1 + n_0$ so that there are n individuals altogether. For $i = 1, ..., n$, let X_i denote the observed vector of covariates and D_i the binary variable indicating participation; $D_i = 1$ if individual i belongs to the treatment group and $D_i = 0$ if individual i belongs to the control group. Let

$$Y_i = Y_{1i} D_i + Y_{0i}(1 - D_i)$$

denote the observed outcome for individual i. We have a random sample $\{Y_i, X_i, D_i\}_{i=1}^n$. In the literature on program evaluation with selection-on-observables, the following two assumptions are often used to evaluate the effect of treatment or program, see e.g., Rosenbaum and Rubin (1983a,b), Hahn (1998), Heckman, Ichimura, Smith, and Todd (1998), Dehejia and Wahba (1999), and Hirano, Imbens, and Ridder (2000), to name only a few.

(C1) Let (Y_1, Y_0, D, X) have a joint distribution. For all $x \in \mathcal{X}$ (the support of X), (Y_1, Y_0) is jointly independent of D conditional on $X = x$.

(C2) For all $x \in \mathcal{X}$, $0 < p(x) < 1$, where $p(x) = P(D = 1|x)$.

In the following, we present sharp bounds on the distribution of Δ under (C1) and (C2). For any fixed $x \in \mathcal{X}$, Lemma 1 provides sharp bounds on the conditional distribution of Δ given $X = x$:

$$F^L(\delta|x) \leq F_\Delta(\delta|x) \leq F^U(\delta|x),$$

where

$$\begin{aligned} F^L(\delta|x) &= \sup_y \max(F_1(y|x) - F_0(y - \delta|x), 0), \\ F^U(\delta|x) &= 1 + \inf_y \min(F_1(y|x) - F_0(y - \delta|x), 0). \end{aligned}$$

Here, we use $F_\Delta(\cdot|x)$ to denote the conditional distribution function of Δ given $X = x$. The other conditional distributions are defined similarly. Conditions (C1) and (C2) allow the identification of the conditional distributions $F_1(y|x)$ and $F_0(y|x)$ appearing in the sharp bounds on $F_\Delta(\delta|x)$. To see this, note that

$$\begin{aligned} F_1(y|x) &= P(Y_1 \leq y|X = x) = P(Y_1 \leq y|X = x, D = 1) \\ &= P(Y \leq y|X = x, D = 1), \end{aligned} \tag{III.15}$$

where (C1) is used to establish the second equality. Similarly, we get

$$F_0(y|x) = P(Y \leq y|X = x, D = 0). \tag{III.16}$$

Given the random sample $\{Y_i, X_i, D_i\}_{i=1}^n$, nonparametric estimators of the bounds $F^L(\delta|x), F^U(\delta|x)$ can be easily constructed from nonparametric estimators of $F_1(y_1|x)$ and $F_0(y_0|x)$. Their asymptotic properties extend directly from those of $F^L(\delta), F^U(\delta)$ established in Section III.3.

Sharp bounds on the unconditional distribution of Δ follow from those of the conditional distribution:

$$E\left(F^L(\delta|X)\right) \leq F_\Delta(\delta) = E\left(F_\Delta(\delta|X)\right) \leq E\left(F^U(\delta|X)\right).$$

We note that if X is independent of (Y_1, Y_0), then the above bounds on $F_\Delta(\delta)$ reduce to those in Lemma 1. In general, X is not independent of (Y_1, Y_0) and the above bounds are tighter than those in Lemma 1.

Let $\hat{F}_1(y_1|x)$ and $\hat{F}_0(y_0|x)$ denote nonparametric estimators of $F_1(y_1|x)$ and $F_0(y_0|x)$ respectively. The bounds $E\left(F^L(\delta|X)\right), E\left(F^U(\delta|X)\right)$ can be estimated respectively by

$$\frac{1}{n}\sum_{i=1}^{n} \max\left(\sup_y \left\{\hat{F}_1(y|X_i) - \hat{F}_0(y-\delta|X_i)\right\}, 0\right)$$

and

$$1 + \frac{1}{n}\sum_{i=1}^{n} \min\left(\inf_y \left\{\hat{F}_1(y|X_i) - \hat{F}_0(y-\delta|X_i)\right\}, 0\right).$$

For the sake of space, we will present a complete asymptotic theory for these estimators in a separate paper.

Conclusion and Extensions

This chapter is the first to study nonparametric estimation and inference for sharp bounds on the distribution of a difference between two random variables. In the context of program evaluation or evaluation of a binary treatment, the difference between the two potential outcomes measures the program effect or effect of the treatment and hence plays an important role. We have also extended our results to a ratio of two random variables, a measure of the relative treatment effects. As we mentioned in the Introduction, sharp bounds on the distribution of a sum are important in finance and risk management. The results developed in this chapter are directly applicable to a sum of two random variables by redefining the second random variable.

Much work remains to be done. In terms of the sharp bounds, those in this chapter do not make use of any prior information on the possible dependence between the potential outcomes. When such information is available, these bounds can be tightened. In a different paper, we explore sharp bounds taking account of dependence information such as values of dependence measures of the potential outcomes. The focus on randomized experiments in this chapter allows the identification of the marginal distributions. In cases where the marginal distributions themselves are not identifiable but bounds on them can be placed (see, e.g., Manski (1994, 2003), Manski and Pepper (2000), Shaikh and Vytlacil (2005), Blundell, Gosling, Ichimura, and Meghir (2006), Honore and Lleras-Muney (2006)), we can also place bounds on the treatment effect distribution.

In terms of statistical inference, this chapter looked at inference on the sharp bounds themselves. The lower and upper bounds represent respectively the minimum and maximum probabilities that the treatment effects do not exceed a given value. Inference on them should be useful on its own right. Alternatively, as initiated in Horowitz and Manski (2000) and Imbens and Manski (2004), followed by Chernozhukov, Hong, and Tamer (2007), and Romano and Shaikh (2006), among others, one may construct confidence sets for the identified set or the true distribution instead of its bounds. However, existing confidence sets may not directly apply to our context, as the asymptotic distributions of the bounds themselves are discontinuous functions of the model parameters. One of the authors is currently investigating this issue by using the general approach developed in Andrews and Guggenberger (2005a, b).

Another extension that we are currently exploring is to relax the assumption that $y_{\sup,\delta}$ ($y_{\inf,\delta}$) is a unique interior solution. As we demonstrated in Example 2, when the supports of F_1 and F_0 are compact, there are often boundary solutions, i.e., $y_{\sup,\delta}$ or $y_{\inf,\delta}$ lie on the boundary of $\mathcal{Y}_\delta$. Moreover, it is also possible to have multiple solutions for $y_{\sup,\delta}$ and $y_{\inf,\delta}$, some in the interior and some on the boundary. The asymptotic distributions of $F_n^L(\delta)$ and $F_n^U(\delta)$ allowing for these possibilities will be much more complicated than those in Theorem 3.2. We will report inference for this general case in a separate paper.

Appendix A. Technical Proofs

Proof of Proposition 1: Since the proofs of (i) and (ii) are similar, we provide a proof for (i) only. Let

$$Q_n(y,\delta) = F_{1n}(y) - F_{0n}(y-\delta),\ Q(y,\delta) = F_1(y) - F_0(y-\delta).$$

Define

$$\hat{y}_{\sup,\delta} = \arg\sup_y Q_n(y,\delta).$$

Then $M_n(\delta) = Q_n(\hat{y}_{\sup,\delta},\delta)$ and $M(\delta) = Q(y_{\sup,\delta},\delta)$. Let $\overline{M}_n(\delta) = Q_n(y_{\sup,\delta},\delta)$. Obviously, $\sqrt{n_1}\left(\overline{M}_n(\delta) - M(\delta)\right) \Longrightarrow N(0,\sigma_L^2)$. We will complete the proof of (i) in three steps:

1. We show that $\hat{y}_{\sup,\delta} - y_{\sup,\delta} = o_p(1)$;
2. We show that $\hat{y}_{\sup,\delta} - y_{\sup,\delta} = O_p(n_1^{-1/3})$;
3. $\sqrt{n_1}\left(M_n(\delta) - M(\delta)\right)$ has the same limiting distribution as $\sqrt{n_1}\left(\overline{M}_n(\delta) - M(\delta)\right)$.

Proof of 1. By the classical Glivenko-Cantelli theorem, the sequences $\sup_y |F_{1n}(y) - F_1(y)|$ and $\sup_y |F_{0n}(y-\delta) - F_0(y-\delta)|$ converge in probability to zero. Consequently, the sequence $\sup_y |[F_{1n}(y) - F_{0n}(y-\delta)] - [F_1(y) - F_0(y-\delta)]\,|$ also converges in probability to zero. This and A3(i) imply that the sequence $\hat{y}_{\sup,\delta}$ converges in probability to $y_{\sup,\delta}$, see e.g., Theorem 5.7 in van der Vaart (1998).

Proof of 2. We use Theorem 3.2.5 in van der Vaart and Wellner (1996) to establish the rate of convergence for $\hat{y}_{\sup,\delta}$. Given (A2), the map: $y \mapsto Q(y,\delta)$ is twice differentiable and has a maximum at $y_{\sup,\delta}$. By (A3), the first condition of Theorem 3.2.5 in van der Vaart and Wellner (1996) is satisfied with $\alpha = 2$. To check the second condition of Theorem 3.2.5 in van der Vaart and Wellner (1996), we consider the centered process:

$$\begin{aligned}\sqrt{n_1}(Q_n - Q)(\cdot,\delta) &= \sqrt{n_1}(F_{1n} - F_1)(\cdot) - \sqrt{n_1}(F_{0n} - F_0)(\cdot - \delta) \\ &\equiv G_{n1}(\cdot) - \frac{\sqrt{n_1}}{\sqrt{n_0}} G_{n0}(\cdot - \delta).\end{aligned}$$

For any $\eta > 0$,

$$\begin{aligned}&E \sup_{|y - y_{\sup,\delta}| < \eta} |\sqrt{n_1}(Q_n - Q)(y,\delta) - \sqrt{n_1}(Q_n - Q)(y_{\sup,\delta},\delta)| \\ &\leq E \sup_{|y - y_{\sup,\delta}| < \eta} |G_{n1}(y) - G_{n1}(y_{\sup,\delta})| \\ &\quad + \frac{\sqrt{n_0}}{\sqrt{n_1}} E \sup_{|y - y_{\sup,\delta}| < \eta} |G_{n0}(y - \delta) - G_{n0}(y_{\sup,\delta} - \delta)|.\end{aligned}$$

Note that the envelope function of the class of functions

$$\{I\{(-\infty, y]\} - I\{(-\infty, y_{\sup,\delta}\} : y \in [y_{\sup,\delta} - \eta, y_{\sup,\delta} + \eta]\}$$

is bounded by $I\{(y_{\sup,\delta} - \eta, y_{\sup,\delta} + \eta)\}$ which has a squared L_2-norm bounded by $2\left[\sup_y f_1(y)\right]\eta$. Since the class of functions $I\{Y_{1i} \leq \cdot\}$ has a finite uniform entropy integral, Lemma 19.38 in van der Vaart (1998) implies:

$$E \sup_{|y - y_{\sup,\delta}| < \eta} |G_{n1}(y) - G_{n1}(y_{\sup,\delta})| \lesssim \eta^{1/2}. \tag{III.17}$$

Similarly, we can show that

$$E \sup_{|y - y_{\sup,\delta}| < \eta} |G_{n0}(y - \delta) - G_{n0}(y_{\sup,\delta} - \delta)| \lesssim \eta^{1/2}. \tag{III.18}$$

Consequently,

$$E \sup_{|y - y_{\sup,\delta}| < \eta} |\sqrt{n_1}(Q_n - Q)(y,\delta) - \sqrt{n_1}(Q_n - Q)(y_{\sup,\delta},\delta)| \lesssim \eta^{1/2}.$$

Hence the second condition of Theorem 3.2.5 in van der Vaart and Wellner (1996) is satisfied leading to the rate of $n_1^{-1/3}$.

Proof of 3. For a fixed δ, we get

$$
\begin{aligned}
&\sqrt{n_1}\left(M_n(\delta)-M(\delta)\right)\\
&= \sqrt{n_1}\left(F_{1n}(\hat{y}_{\sup,\delta})-F_{0n}(\hat{y}_{\sup,\delta}-\delta)\right)-\sqrt{n_1}\left(F_1(y_{\sup,\delta})-F_0(y_{\sup,\delta}-\delta)\right)\\
&= \sqrt{n_1}(Q_n-Q)(\hat{y}_{\sup,\delta},\delta)+\sqrt{n_1}\left(F_1(\hat{y}_{\sup,\delta})-F_0(\hat{y}_{\sup,\delta}-\delta)\right)\\
&\quad -\sqrt{n_1}\left(F_1(y_{\sup,\delta})-F_0(y_{\sup,\delta}-\delta)\right)\\
&= \sqrt{n_1}(Q_n-Q)(y_{\sup,\delta},\delta)\\
&\quad +\sqrt{n_1}\left[F_1(\hat{y}_{\sup,\delta})-F_0(\hat{y}_{\sup,\delta}-\delta)-F_1(y_{\sup,\delta})-F_0(y_{\sup,\delta}-\delta)\right]+o_p(1)\\
&= \sqrt{n_1}\left(\overline{M}_n(\delta)-M(\delta)\right)\\
&\quad +\frac{1}{2}\sqrt{n_1}\left\{f_1'(y^*_{\sup,\delta})-f_0'(y^*_{\sup,\delta}-\delta)\right\}(\hat{y}_{\sup,\delta}-y_{\sup,\delta})^2+o_p(1)\\
&= \sqrt{n_1}\left(\overline{M}_n(\delta)-M(\delta)\right)+o_p(1),
\end{aligned}
$$

where $y^*_{\sup,\delta}$ lies between $\hat{y}_{\sup,\delta}$ and $y_{\sup,\delta}$ and we have used stochastic equicontinuity of the process: $\sqrt{n_1}(Q_n-Q)(\cdot,\delta)$ and the first order condition for $\sup_y\{F_1(y)-F_0(y-\delta)\}$.

Appendix B. Functional Forms of $y_{\sup,\delta}$, $y_{\inf,\delta}$, $M(\delta)$ and $m(\delta)$ for Some Known Marginal Distributions

Denuit, Genest, and Marceau (1999) provided the distribution bounds for a sum of two random variables when they both follow shifted Exponential distributions or both follow shifted Pareto distributions. Below, we augment their results with explicit expressions for $y_{\sup,\delta}$, $y_{\inf,\delta}$, $M(\delta)$ and $m(\delta)$ which may help us understand the asymptotic behavior of the nonparametric estimators of the distribution bounds when the true marginals are either shifted Exponential or shifted Pareto.

First, we present some expressions used in Example 2.

Example 2 (continued). In Example 2, we considered the family of distributions denoted by $C(a)$ with $a \in (0,1)$. If $X \sim C(a)$, then

$$F(x) = \begin{cases} \dfrac{1}{a}x^2 & \text{if } x \in [0,a] \\ 1 - \dfrac{(x-1)^2}{(1-a)} & \text{if } x \in [a,1] \end{cases} \quad \text{and}$$

$$f(x) = \begin{cases} \dfrac{2}{a}x & \text{if } x \in [0,a] \\ \dfrac{2(1-x)}{(1-a)} & \text{if } x \in [a,1] \end{cases}.$$

Suppose $Y_1 \sim C(a_1)$ and $Y_0 \sim C(a_0)$. We now provide the functional form of $F_1(y) - F_0(y-\delta)$.

1. Suppose $\delta < 0$. Then $\mathcal{Y}_\delta = [0, 1+\delta]$.

   (a) If $a_0 + \delta \leq 0 < a_1 \leq 1+\delta$, then

$$F_1(y) - F_0(y-\delta) = \begin{cases} \frac{y^2}{a_1} - \left(1 - \frac{(y-\delta-1)^2}{(1-a_0)}\right) & \text{if } 0 \leq y \leq a_1 \\ \left(1 - \frac{(y-1)^2}{(1-a_1)}\right) - \left(1 - \frac{(y-\delta-1)^2}{(1-a_0)}\right) & \text{if } a_1 \leq y \leq 1+\delta \end{cases};$$

(b) If $0 \leq a_0 + \delta \leq a_1 \leq 1 + \delta$, then

$$F_1(y) - F_0(y-\delta) = \begin{cases} \frac{y^2}{a_1} - \frac{(y-\delta)^2}{a_0} \text{ if } 0 \leq y \leq a_0 + \delta \\ \frac{y^2}{a_1} - \left(1 - \frac{(y-\delta-1)^2}{(1-a_0)}\right) \text{ if } a_0 + \delta \leq y \leq a_1 \\ \left(1 - \frac{(y-1)^2}{(1-a_1)}\right) - \left(1 - \frac{(y-\delta-1)^2}{(1-a_0)}\right) \text{ if } a_1 \leq y \leq 1 + \delta \end{cases};$$

(c) If $a_0 + \delta \leq 0 \leq 1 + \delta \leq a_1$, then

$$F_1(y) - F_0(y-\delta) = \frac{y^2}{a_1} - \left(1 - \frac{(y-\delta-1)^2}{(1-a_0)}\right) \text{ if } 0 \leq y \leq 1 + \delta;$$

(d) If $0 \leq a_0 + \delta < 1 + \delta \leq a_1$, then

$$F_1(y) - F_0(y-\delta) = \begin{cases} \frac{y^2}{a_1} - \frac{(y-\delta)^2}{a_0} \text{ if } 0 \leq y \leq a_0 + \delta \\ \frac{y^2}{a_1} - \left(1 - \frac{(y-\delta-1)^2}{(1-a_0)}\right) \text{ if } a_0 + \delta \leq y \leq 1 + \delta \end{cases};$$

(e) If $0 < a_1 \leq a_0 + \delta \leq 1 + \delta$, then

$$F_1(y) - F_0(y-\delta) = \begin{cases} \frac{y^2}{a_1} - \frac{(y-\delta)^2}{a_0} \text{ if } 0 \leq y \leq a_1 \\ \left(1 - \frac{(y-1)^2}{(1-a_1)}\right) - \frac{(y-\delta)^2}{a_0} \text{ if } a_1 \leq y \leq a_0 + \delta \\ \left(1 - \frac{(y-1)^2}{(1-a_1)}\right) - \left(1 - \frac{(y-\delta-1)^2}{(1-a_0)}\right) \text{ if } a_0 + \delta \leq y \leq 1 + \delta \end{cases}.$$

2. Suppose $\delta \geq 0$. Then $\mathcal{Y}_\delta = [\delta, 1]$.

(a) If $\delta < a_0 + \delta \leq a_1 < 1$, then

(i) if $a_1 \neq a_0$ and $\delta \neq 0$, then

$$F_1(y) - F_0(y-\delta) = \begin{cases} \frac{y^2}{a_1} - \frac{(y-\delta)^2}{a_0} \text{ if } \delta \leq y \leq a_0 + \delta \\ \frac{y^2}{a_1} - \left(1 - \frac{(y-\delta-1)^2}{(1-a_0)}\right) \text{ if } a_0 + \delta \leq y \leq a_1 \\ \left(1 - \frac{(y-1)^2}{(1-a_1)}\right) - \left(1 - \frac{(y-\delta-1)^2}{(1-a_0)}\right) \text{ if } a_1 \leq y \leq 1 \end{cases};$$

(ii) if $a_1 = a_0 = a$ and $\delta = 0$, then

$$F_1(y) - F_0(y-\delta) = 0 \text{ for all } y \in [0,1].$$

(b) If $\delta \leq a_1 \leq a_0 + \delta \leq 1$, then

$$F_1(y) - F_0(y-\delta) = \begin{cases} \frac{y^2}{a_1} - \frac{(y-\delta)^2}{a_0} \text{ if } \delta \leq y \leq a_1 \\ \left(1 - \frac{(y-1)^2}{(1-a_1)}\right) - \frac{(y-\delta)^2}{a_0} \text{ if } a_1 \leq y \leq a_0 + \delta \\ \left(1 - \frac{(y-1)^2}{(1-a_1)}\right) - \left(1 - \frac{(y-\delta-1)^2}{(1-a_0)}\right) \text{ if } a_0 + \delta \leq y \leq 1 \end{cases};$$

(c) If $\delta \leq a_1 < 1 \leq a_0 + \delta$, then

$$F_1(y) - F_0(y-\delta) = \begin{cases} \frac{y^2}{a_1} - \frac{(y-\delta)^2}{a_0} \text{ if } \delta \leq y \leq a_1 \\ \left(1 - \frac{(y-1)^2}{(1-a_1)}\right) - \frac{(y-\delta)^2}{a_0} \text{ if } a_1 \leq y \leq 1 \end{cases};$$

(d) If $a_1 < \delta < a_0 + \delta \leq 1$, then

$$F_1(y) - F_0(y-\delta) = \begin{cases} \left(1 - \frac{(y-1)^2}{(1-a_1)}\right) - \frac{(y-\delta)^2}{a_0} \text{ if } \delta \leq y \leq a_0 + \delta \\ \left(1 - \frac{(y-1)^2}{(1-a_1)}\right) - \left(1 - \frac{(y-\delta-1)^2}{(1-a_0)}\right) \text{ if } a_0 + \delta \leq y \leq 1 \end{cases};$$

(e) If $a_1 < \delta < 1 \leq a_0 + \delta$, then

$$F_1(y) - F_0(y-\delta) = \left(1 - \frac{(y-1)^2}{(1-a_1)}\right) - \frac{(y-\delta)^2}{a_0} if \delta \leq y \leq 1.$$

(Shifted) Exponential marginals. The marginal distributions are:

$$\begin{aligned} F_1(y) &= 1 - \exp\left(-\tfrac{y-\theta_1}{\alpha_1}\right) \text{ for } y \in [\theta_1, \infty) \text{ and} \\ F_0(y) &= 1 - \exp\left(-\tfrac{y-\theta_0}{\alpha_0}\right) \text{ for } y \in [\theta_0, \infty)\text{, where } \alpha_1, \theta_1, \alpha_0, \theta_0 > 0. \end{aligned}$$

Let $\delta_c = (\theta_1 - \theta_0) - \min\{\alpha_1, \alpha_0\}(\ln \alpha_1 - \ln \alpha_0)$.

1. Suppose $\alpha_1 < \alpha_0$.

(a) If $\delta \leq \delta_c$,

$$\begin{aligned}
F^L(\delta) &= \max\{M(\delta), 0\} = 0, \\
\text{where } M(\delta) &= \left(\left(\frac{\alpha_0}{\alpha_1}\right)^{\frac{\alpha_1}{\alpha_1-\alpha_0}} - \left(\frac{\alpha_0}{\alpha_1}\right)^{\frac{\alpha_0}{\alpha_1-\alpha_0}}\right)\exp\left(-\frac{\delta-(\theta_1-\theta_0)}{\alpha_1-\alpha_0}\right) < 0, \\
\text{and } y_{\inf,\delta} &= \frac{\alpha_0\alpha_1(\ln\alpha_1 - \ln\alpha_0) + \alpha_1\theta_0 - \alpha_0\theta_1 + \alpha_1\delta}{\alpha_1-\alpha_0} \text{ (an interior solution).} \\
F^U(\delta) &= 1 + \min\{m(\delta), 0\} = 1 + m(\delta), \\
\text{where } m(\delta) &= \min\left\{\exp\left(-\tfrac{\max\{\theta_1-(\delta+\theta_0),0\}}{\alpha_0}\right) - \exp\left(-\tfrac{\max\{\theta_0+\delta-\theta_1,0\}}{\alpha_1}\right), 0\right\} \\
\text{and } y_{\sup,\delta} &= \max\{\theta_1, \theta_0+\delta\} \text{ or } \infty \text{ (boundary solution).}
\end{aligned}$$

(b) If $\delta > \delta_c$,

$$\begin{aligned}
F^L(\delta) &= \max\{M(\delta), 0\} = M(\delta) > 0, \\
\text{where } M(\delta) &= 1 - \exp\left(-\tfrac{\delta+\theta_0-\theta_1}{\alpha_1}\right) \text{ and } y_{\inf,\delta} = \theta_0 + \delta. \\
F^U(\delta) &= 1 + \min\{m(\delta), 0\} = 1 \\
\text{since } m(\delta) &= 0 \text{ and } y_{\sup,\delta} = \infty.
\end{aligned}$$

2. Suppose $\alpha_1 = \alpha_0 = \alpha$. Then

$$
\begin{aligned}
F^L(\delta) &= \max\{M(\delta), 0\} = M(\delta), \\
\text{where } M(\delta) &= \begin{cases} 0 \text{ if } \delta \leq \theta_1 - \theta_0 \\ 1 - \exp\left(-\frac{\delta - (\theta_1 - \theta_0)}{\alpha}\right) > 0 \text{ if } \delta > \theta_1 - \theta_0 \end{cases} \\
\text{and } y_{\inf,\delta} &= \begin{cases} \infty \text{ if } \delta < \theta_1 - \theta_0 \\ \text{any point in } \mathcal{R} \text{ if } \delta = \theta_1 - \theta_0 \\ \theta_0 + \delta \text{ if } \delta > \theta_1 - \theta_0 \end{cases} \cdot \\
F^U(\delta) &= 1 + \min\{m(\delta), 0\} = 1 + m(\delta), \\
\text{where } m(\delta) &= \begin{cases} \exp\left(-\frac{\theta_1 - (\delta + \theta_0)}{\alpha}\right) - 1 < 0 \text{ if } \delta < \theta_1 - \theta_0 \\ 0 \text{ if } \delta \geq \theta_1 - \theta_0 \end{cases} \\
\text{and } y_{\sup,\delta} &= \begin{cases} \theta_1 \text{ if } \delta < \theta_1 - \theta_0 \\ \text{any point in } \mathcal{R} \text{ if } \delta = \theta_1 - \theta_0 \\ \infty \text{ if } \delta > \theta_1 - \theta_0 \end{cases} \cdot
\end{aligned}
$$

3. Suppose $\alpha_1 > \alpha_0$.

   (a) If $\delta < \delta_c$,

$$
\begin{aligned}
F^L(\delta) &= \max\{M(\delta), 0\} = 0 \text{ , since } M(\delta) = 0 \text{ and } y_{\inf,\delta} = \infty. \\
F^U(\delta) &= 1 + \min\{m(\delta), 0\} = 1 - m(\delta), \\
\text{where } m(\delta) &= \exp\left(-\frac{\theta_1 - (\delta + \theta_0)}{\alpha_0}\right) - 1 < 0, \;\; y_{\sup,\delta} = \theta_1.
\end{aligned}
$$

(b) If $\delta \geq \delta_c$,

$$\begin{aligned}
F^L(\delta) &= \max\{M(\delta), 0\} = M(\delta), \\
\text{where } M(\delta) &= \max\left\{\exp\left(-\tfrac{\max\{\theta_1-(\delta+\theta_0),0\}}{\alpha_0}\right) - \exp\left(-\tfrac{\max\{\theta_0+\delta-\theta_1,0\}}{\alpha_1}\right), 0\right\} \\
\text{and } y_{\inf,\delta} &= \max\{\theta_1, \theta_0+\delta\} \text{ or } \infty \text{ (boundary solution)}. \\
F^U(\delta) &= 1+\min\{m(\delta), 0\} = 1+m(\delta), \\
\text{where } m(\delta) &= \left(\left(\frac{\alpha_0}{\alpha_1}\right)^{\frac{\alpha_1}{\alpha_1-\alpha_0}} - \left(\frac{\alpha_0}{\alpha_1}\right)^{\frac{\alpha_0}{\alpha_1-\alpha_0}}\right)\exp\left(-\frac{\delta-(\theta_1-\theta_0)}{\alpha_1-\alpha_0}\right) < 0 \\
\text{and } y_{\sup,\delta} &= \frac{\alpha_0\alpha_1(\ln\alpha_1 - \ln\alpha_0) + \alpha_1\theta_0 - \alpha_0\theta_1 + \alpha_1\delta}{\alpha_1-\alpha_0} \text{ (an interior solution).}
\end{aligned}$$

(Shifted) Pareto marginals. The marginal distributions are:

$$\begin{aligned}
F_1(y) &= 1-\left(\frac{\lambda_1}{\lambda_1+y-\theta_1}\right)^{\alpha} \text{ for } y \in [\theta_1, \infty) \text{ and} \\
F_0(y) &= 1-\left(\frac{\lambda_0}{\lambda_0+y-\theta_0}\right)^{\alpha} \text{ for } y \in [\theta_0, \infty), \text{ where } \alpha, \lambda_1, \theta_1, \lambda_0, \theta_0 > 0.
\end{aligned}$$

Define

$$\delta_c = (\theta_1-\theta_0) - (\max\{\lambda_1, \lambda_0\})^{\frac{\alpha}{\alpha+1}}\left(\lambda_1^{\frac{1}{\alpha+1}} - \lambda_0^{\frac{1}{\alpha+1}}\right).$$

1. Suppose $\lambda_1 < \lambda_0$.

(a) If $\delta \leq \delta_c$, then

$$
\begin{aligned}
F^L(\delta) &= \max\{M(\delta), 0\} = M(\delta), \\
\text{where } M(\delta) &= \left(\lambda_0^{\frac{\alpha}{\alpha+1}} - \lambda_1^{\frac{\alpha}{\alpha+1}}\right)\left(\frac{\lambda_1^{\frac{\alpha}{\alpha+1}} - \lambda_0^{\frac{\alpha}{\alpha+1}}}{\delta - \lambda_0 + \lambda_1 - \theta_1 + \theta_0}\right)^{\alpha} > 0 \\
\text{and } y_{\inf,\delta} &= \frac{(\delta + \theta_0 - \lambda_0)\lambda_1^{\frac{\alpha}{\alpha+1}} + (\lambda_1 - \theta_1)\lambda_0^{\frac{\alpha}{\alpha+1}}}{\lambda_1^{\frac{\alpha}{\alpha+1}} - \lambda_0^{\frac{\alpha}{\alpha+1}}} \text{ (an interior solution).} \\
F^U(\delta) &= 1 + \min\{m(\delta), 0\} = 1 + m(\delta), \\
\text{where } m(\delta) &= \min\left\{\left(\frac{\lambda_0}{\lambda_0 + \max\{\theta_1 - \delta - \theta_0, 0\}}\right)^{\alpha}\right. \\
&\quad \left. - \left(\frac{\lambda_1}{\lambda_1 + \max\{\theta_0 + \delta - \theta_1, 0\}}\right)^{\alpha}, 0\right\} \\
\text{and } y_{\sup,\delta} &= \max\{\theta_1, \theta_0 + \delta\} \text{ or } \infty \text{ (boundary solution).}
\end{aligned}
$$

(b) If $\delta > \delta_c$, then

$$
\begin{aligned}
F^L(\delta) &= \max\{M(\delta), 0\} = M(\delta), \\
\text{where } M(\delta) &= 1 - \left(\frac{\lambda_1}{\lambda_1 + \theta_0 + \delta - \theta_1}\right)^{\alpha} \geq 0 \text{ and } y_{\inf,\delta} = \theta_0 + \delta. \\
F^U(\delta) &= 1 + \min\{m(\delta), 0\} = 1 \\
\text{since } m(\delta) &= 0 \text{ and } y_{\sup,\delta} = \infty.
\end{aligned}
$$

2. Suppose $\lambda_1 = \lambda_0 = \lambda$. Then

$$\begin{aligned}
F^L(\delta) &= \max\{M(\delta), 0\} = M(\delta), \\
\text{where } M(\delta) &= \begin{cases} 0 \text{ if } \delta \leq \theta_1 - \theta_0 \\ 1 - \left(\dfrac{\lambda}{\lambda + \delta - (\theta_1 - \theta_0)}\right)^{\alpha} \geq 0 \text{ otherwise} \end{cases} \\
\text{and } y_{\inf,\delta} &= \begin{cases} \infty \text{ if } \delta < \theta_1 - \theta_0 \\ \text{any point in } \mathcal{Y} \text{ if } \delta = \theta_1 - \theta_0 \\ \theta_0 + \delta \text{ if } \delta > \theta_1 - \theta_0 \end{cases} . \\
F^U(\delta) &= 1 + \min\{m(\delta), 0\} = 1 + m(\delta), \\
\text{where } m(\delta) &= \begin{cases} \left(\dfrac{\lambda}{\lambda - \delta + (\theta_1 - \theta_0)}\right)^{\alpha} - 1 \text{ if } \delta < \theta_1 - \theta_0 \\ 0 \text{ if } \delta \geq \theta_1 - \theta_0 \end{cases} \\
\text{and } y_{\sup,\delta} &= \begin{cases} \theta_1 \text{ if } \delta < \theta_1 - \theta_0 \\ \text{any point in } \mathcal{Y} \text{ if } \delta = \theta_1 - \theta_0 \\ \infty \text{ if } \delta > \theta_1 - \theta_0 \end{cases} .
\end{aligned}$$

3. Suppose $\lambda_1 > \lambda_0$.

   (a) If $\delta < \delta_c$, then

$$\begin{aligned}
F^L(\delta) &= \max\{M(\delta), 0\} = 0 \text{ since } M(\delta) = 0 \text{ and} \\
y_{\inf,\delta} &= \infty. \\
F^U(\delta) &= 1 + \min\{m(\delta), 0\} = 1 + m(\delta), \\
\text{where } m(\delta) &= \left(\frac{\lambda_0}{\lambda_0 + \theta_1 - \delta - \theta_0}\right)^{\alpha} - 1 \leq 0 \text{ and} \\
y_{\sup,\delta} &= \theta_1.
\end{aligned}$$

(b) If $\delta \geq \delta_c$, then

$$\begin{aligned}
F^L(\delta) &= \max\{M(\delta), 0\} = M(\delta), \\
\text{where } M(\delta) &= \max\left\{ \left(\frac{\lambda_0}{\lambda_0 + \max\{\theta_1 - \delta - \theta_0, 0\}} \right)^\alpha \right. \\
&\quad \left. - \left(\frac{\lambda_1}{\lambda_1 + \max\{\theta_0 + \delta - \theta_1, 0\}} \right)^\alpha, 0 \right\} \\
\text{and } y_{\inf,\delta} &= \max\{\theta_1, \theta_0 + \delta\} \text{ or } \infty \text{ (boundary solution).}
\end{aligned}$$

$$\begin{aligned}
F^U(\delta) &= 1 + \min\{m(\delta), 0\} = 1 + m(\delta), \\
\text{where } m(\delta) &= \left(\lambda_0^{\frac{\alpha}{\alpha+1}} - \lambda_1^{\frac{\alpha}{\alpha+1}} \right) \left(\frac{\lambda_1^{\frac{\alpha}{\alpha+1}} - \lambda_0^{\frac{\alpha}{\alpha+1}}}{\delta - \lambda_0 + \lambda_1 - \theta_1 + \theta_0} \right)^\alpha < 0 \\
\text{and } y_{\sup,\delta} &= \frac{(\delta + \theta_0 - \lambda_0)\, \lambda_1^{\frac{\alpha}{\alpha+1}} + (\lambda_1 - \theta_1)\, \lambda_0^{\frac{\alpha}{\alpha+1}}}{\lambda_1^{\frac{\alpha}{\alpha+1}} - \lambda_0^{\frac{\alpha}{\alpha+1}}} \text{ (an interior solution).}
\end{aligned}$$

CHAPTER IV

CONFIDENCE SETS FOR THE QUANTILE OF TREATMENT EFFECTS

Introduction

In evaluating the effect of a treatment such as a social program implementation, the gain or loss in the outcome from the treatment or the amount of treatment effect may differ across individuals in the presence of heterogeneous response to treatment among individuals.[1] Researchers take the heterogeneity into consideration by estimating, for instance, the *Average Treatment Effect* (ATE) conditional upon some observable covariates. However, if there exists heterogeneity due to unobservable characteristics, the amount of gain from the treatment may differ across those with the same observable covariates. In this case, the distributional effect of a treatment may not be fully identified even with the conditional ATE on the observable covariates.

In this chapter, we focus on a binary treatment and define the individual treatment effect as the difference between the two potential outcomes: the individual's potential outcome when he/she is assigned to the treated group and the potential outcome when he/she is assigned to the control group. Given that only one of the two potential outcomes is observed for any individual, we cannot observe the individual's outcome gain. This missing value problem is the fundamental obstacle to the identification of the distribution of the treatment effects when there is unobserved heterogeneity. Without imposing strong dependence structure on the potential outcomes, Chapter III investigated the (partial) identification

[1] We are going to use 'outcome gain' and 'treatment effect' interchangably throughout the rest of this paper.

of the distribution of the effects of a binary treatment in a randomized experiment. In a randomized experiment, the marginal distributions of the potential outcomes are identified. Given the marginal distributions, the distribution of the treatment effects is partially (interval) identified. Chapter III established asymptotic properties of nonparametric estimators of the bounds of the identified interval and provided valid inference procedures for the true bounds.

In this chapter, we take a different approach to tackle the heterogeneity of treatment effects. Differently from Chapter III, we will focus on the quantiles of the treatment effects denoted as $Q_{\mathrm{TE}}(p)$ for quantile level $0 < p < 1$. The $Q_{\mathrm{TE}}(p)$ is of interest in various situations. For example, policy makers may want to know if the median outcome gain ($p = 1/2$) is positive or not. When we investigate the changes of income due to a program, we want to trace out the changes of income of a certain quantile. Various inequality measures are functions of quantiles.

The importance of the $Q_{\mathrm{TE}}(p)$ in capturing the heterogeneity of treatment effects has been recognized in the literature. Most existing literature, however, focus on the *Quantile Treatment Effects* (QTE) - the difference between the quantiles of the outcomes of the treated and control groups at a given quantile level p - rather than the $Q_{\mathrm{TE}}(p)$. The QTE has been used as an indicator of the presence of heterogeneity in treatment effects since Lehmann (1974) and Doksum (1974) first introduced the concept. For example, Bitler, Gelbach, and Hoynes (1996) used QTE to show that ATE only is not enough to capture the distributional impact of social programs. Recently, due to rapidly growing interest on the heterogeneity in the treatment effects, QTE has received a great deal of attention. To name only a few: Abadie, Angrist, and Imbens (2002) estimated QTE and investigated the heterogeneity of treatment effects in the training program provided under the Job Training

Partnership Act (JTPA); Djebbari and Smith (2008) showed heterogeneous treatment effects in PROGRESSA by estimating QTE; Millimet, Daniel L. and Abdullah Kumas (2007) studied the heterogeneous impact of taxation on the US' foreign direct investment by investigating QTE; Ma and Koenker (2006) estimated QTE of class size on the performance of Dutch primary school students.

Chapter III showed that the QTE thus defined is the same as the $Q_{\mathrm{TE}}(p)$ only under the assumption that the two potential outcomes are perfectly positively dependent (Firpo (2007) called this a *rank preservation* assumption) and when the QTE is nondecreasing in the quantile level p. More fundamental difference between QTE and $Q_{\mathrm{TE}}(p)$ is that the QTE, being the difference between the quantiles of the marginal distributions of the potential outcomes, is identified as long as the marginal distributions are identified whereas $Q_{\mathrm{TE}}(p)$ can only be partially identified unless strong dependence structure is imposed on the potential outcomes. Estimation and statistical inference procedures about the QTE have been established (see, for example, Chernozhukov and Hansen 2006 for parametric framework and Firpo 2007 for nonparametric framework) while, to the best of our knowledge, no systematic study on the identification and inference on $Q_{\mathrm{TE}}(p)$ is currently available except for a preliminary attempt to find bounds for the $Q_{\mathrm{TE}}(p)$ by extending the concept of QTE with various correlations between two potential outcomes (Heckman, Smith, and Clements 1997).

In this chapter, we explore the partial identification of $Q_{\mathrm{TE}}(p)$ and provide nonparametric estimators for its identified bounds. We establish asymptotic properties of these estimators and provide confidence intervals (CI) for these bounds and confidence set (CS) for the $Q_{\mathrm{TE}}(p)$ itself. Due to the partial identification of $Q_{\mathrm{TE}}(p)$, we make use of the recent developments in inference procedures for partially identified models to construct CSs for

$Q_{\text{TE}}(p)$.

The rest of this chapter is organized as follows. In Section 2, we introduce sharp bounds for $Q_{\text{TE}}(p)$, provide their nonparametric estimators, and develop the asymptotic theory for these estimators. In contrast to the asymptotic distributions of estimators of the distribution bounds studied in Chapter III, the asymptotic distributions of estimators of the bounds for $Q_{\text{TE}}(p)$ are continuous in model parameters which simplify statistical inference procedures for the bounds and the $Q_{\text{TE}}(p)$. Section 3 presents various confidence sets. For each bound, we propose a new confidence interval which does not require the estimation of the probability density functions of the potential outcomes. In addition to the direct application of the CSs in Chapter II to $Q_{\text{TE}}(p)$, we also present a new CS for the true $Q_{\text{TE}}(p)$ which does not require the estimation of the density functions of the potential outcomes. Monte Carlo Simulation results are presented in Section 4. Section 5 concludes. Technical proofs are gathered in Appendix A. Appendix B presents the graphs of the distribution functions and density functions used to generate the potential outcomes in the simulation study. Appendices C and D present respectively tables of simulation results for the bounds and the $Q_{\text{TE}}(p)$.

Nonparametric Estimators of Sharp Bounds on $Q_{\mathbf{TE}}(p)$ and Their Asymptotic Properties

Sharp Bounds on $Q_{\mathbf{TE}}(p)$ and Their Estimators

The notation in this chapter follows the convention in the treatment effect literature. We consider a binary treatment and use Y_1 to denote the potential outcome from receiving treatment and Y_0 the outcome without treatment. Let $F(y_1, y_0)$ denote the joint distribution of Y_1, Y_0 with marginals $F_1(\cdot)$ and $F_0(\cdot)$ respectively. Let $\Delta = Y_1 - Y_0$ denote the treatment effect or outcome gain and $F_\Delta(\cdot)$ its distribution function. Let $F_\Delta^{-1}(\cdot)$ denote the generalized inverse of $F_\Delta(\cdot)$. For a given quantile level p, $Q_{\text{TE}} = Q_{\text{TE}}(p) = F_\Delta^{-1}(p)$. Given the marginals F_1 and F_0, sharp bounds on the Q_{TE} can be found in Williamson and Downs (1990). They are restated in the following lemma.

Lemma 5 *For $0 < p < 1$, $Q^L(p) \leq Q_{TE}(p) \leq Q^U(p)$, where*

$$\begin{aligned} Q^U(p) &= \inf_{u \in (p,1)} [F_1^{-1}(u) - F_0^{-1}(u-p)], \\ Q^L(p) &= \sup_{u \in (0,p)} [F_1^{-1}(u) - F_0^{-1}(1+u-p)], \end{aligned}$$

and these bounds are sharp.

Example. When $Y_1 \sim N\left(\mu_1, \sigma_1^2\right)$ and $Y_0 \sim N\left(\mu_0, \sigma_0^2\right)$. Provided that $u_{\sup,p} \neq 0.5$ and $1 + u_{\sup,p} - p \neq 0.5$ and that $u_{\inf,p} \neq 0.5$ and $u_{\inf,p} - p \neq 0.5$, define $u_{\sup,p}$ and $u_{\inf,p}$ such as

$$\begin{aligned} u_{\sup,p} &= \arg_{u \in (0,p)} \left\{ \left(\Phi^{-1}(u)\right)^2 - \left(\Phi^{-1}(1+u-p)\right)^2 = \ln \frac{\sigma_0}{\sigma_1} \right\} \text{ for any } p \neq 1, \\ u_{\inf,p} &= \arg_{u \in (p,1)} \left\{ \left(\Phi^{-1}(u)\right)^2 - \left(\Phi^{-1}(u-p)\right)^2 = \ln \frac{\sigma_0}{\sigma_1} \right\} \text{ for any } p \neq 0. \end{aligned}$$

Then,

$$\begin{aligned} Q^U(p) &= \mu_1 - \mu_0 + \sigma_1 \Phi^{-1}(u_{\inf,p}) - \sigma_1 \Phi^{-1}(u_{\inf,p} - p), \\ Q^L(p) &= \mu_1 - \mu_0 + \sigma_1 \Phi^{-1}(u_{\inf,p}) - \sigma_1 \Phi^{-1}(1 + u_{\sup,p} - p). \end{aligned}$$

Suppose random samples $\{Y_{1i}\}_{i=1}^{n_1} \sim F_1$ and $\{Y_{0i}\}_{i=1}^{n_0} \sim F_0$ are available. For a $p \in \left[\frac{1}{\max\{n_1,n_0\}}, 1 - \frac{1}{\max\{n_1,n_0\}}\right]$, we provide the following estimators of $Q^L(p)$ and $Q^U(p)$:

$$Q_n^L(p) = \sup_{u \in \left[\frac{1}{\max\{n_1,n_0\}}, p\right]} \{F_{1n}^{-1}(u) - F_{0n}^{-1}(1+u-p)\}, \tag{IV.1}$$

$$Q_n^U(p) = \inf_{u \in \left[p, 1-\frac{1}{\max\{n_1,n_0\}}\right]} \{F_{1n}^{-1}(u) - F_{0n}^{-1}(u-p)\}, \tag{IV.2}$$

where $F_{1n}(\cdot)$ and $F_{0n}(\cdot)$ are the empirical distributions defined as

$$F_{kn}(y) = \frac{1}{n_k} \sum_{i=1}^{n_k} 1\{Y_{ki} \leq y\}, \; k = 1, 0,$$

and $F_{1n}^{-1}(\cdot)$ and $F_{0n}^{-1}(\cdot)$ are the generalized empirical quantile functions.

For any fixed p, the consistency of $Q_n^L(p)$ and $Q_n^U(p)$ is obvious. In the next subsection, we will establish the asymptotic distributions of

$$\sqrt{n_1} \begin{pmatrix} Q_n^L(p) - Q^L(p) \\ Q_n^U(p) - Q^U(p) \end{pmatrix}.$$

Asymptotic Distributions

We make the following assumptions.

(A1) (i) The two samples $\{Y_{1i}\}_{i=1}^{n_1}$ and $\{Y_{0i}\}_{i=1}^{n_0}$ are each i.i.d. and are independent of each other; (ii) $n_1/n_0 \to \lambda$ as $n_1 \to \infty$ with $0 < \lambda < \infty$.

(A2) (i) The distribution functions F_1 and F_0 are twice differentiable with positive, bounded density functions f_1 and f_0 on their supports;
(ii) $\inf_u f_1\left(F_1^{-1}(u)\right) > 0$ and $\inf_u f_0\left(F_0^{-1}(u)\right) > 0$.

(A3) The function $u \longmapsto \left\{F_1^{-1}(u) - F_0^{-1}(1 + u - p)\right\}$ has a unique maximum at $u_{\sup,p}$ in the interior of $[0, p]$.

(A4) The function $u \longmapsto \left\{F_1^{-1}(u) - F_0^{-1}(u - p)\right\}$ has a unique minimum at $u_{\inf,p}$ in the interior of $[p, 1]$.

(A1) is satisfied in a randomized experiment. Under (A3) and (A4), we can define $u_{\sup,p}$ and $u_{\inf,p}$ as follows:

$$\begin{aligned} u_{\sup,p} &= \arg\sup_{u\in(0,p)} \left\{F_1^{-1}(u) - F_0^{-1}(1+u-p)\right\}, \\ u_{\inf,p} &= \arg\inf_{u\in(p,1)} \left\{F_1^{-1}(u) - F_0^{-1}(u-p)\right\}. \end{aligned}$$

(A3) and (A4) can be restrictive. In general, if both Y_1 and Y_0 have unbounded supports, then the interiority of $u_{\sup,p}$ and $u_{\inf,p}$ is not questionable but the uniqueness is not always guaranteed. If either of them has a bounded support from above, below, or both, then neither the interiority nor the uniqueness is guaranteed. In future work, we will relax (A3) and (A4).

With such $u_{\sup,p}$ and $u_{\inf,p}$, we have:

$$Q^U(p) = F_1^{-1}(u_{\inf,p}) - F_0^{-1}(u_{\inf,p} - p), \tag{IV.3}$$

$$Q^L(p) = F_1^{-1}(u_{\sup,p}) - F_0^{-1}(1 + u_{\sup,p} - p). \tag{IV.4}$$

Theorem 5 below provides the asymptotic distribution of $(Q_n^L(p), Q_n^U(p))$. For the ease of exposition, we define

$$\begin{aligned} \hat{u}_{\sup,p} &= \inf\left\{\arg \sup_{u\in\left[\frac{1}{\max\{n_1,n_0\}},p\right]} \{F_{1n}^{-1}(u) - F_{0n}^{-1}(1+u-p)\}\right\}, \\ \hat{u}_{\inf,p} &= \inf\left\{\arg \inf_{u\in\left[p,1-\frac{1}{\max\{n_1,n_0\}}\right]} \{F_{1n}^{-1}(u) - F_{0n}^{-1}(u-p)\}\right\}. \end{aligned}$$

Then we have

$$Q_n^U(p) = F_{1n}^{-1}(\hat{u}_{\inf,p}) - F_{0n}^{-1}(\hat{u}_{\inf,p} - p),\ Q_n^L(p) = F_{1n}^{-1}(\hat{u}_{\sup,p}) - F_{0n}^{-1}(1 + \hat{u}_{\sup,p} - p).$$

Theorem 5 *Suppose (A1)-(A4) hold. Then*

$$\sqrt{n_1}\begin{pmatrix} Q_n^L(p) - Q^L(p) \\ Q_n^U(p) - Q^U(p) \end{pmatrix} \Rightarrow N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_L^2 & \sigma_{LU} \\ \sigma_{LU} & \sigma_U^2 \end{pmatrix}\right),$$

where

$$\begin{aligned}
\sigma_L^2 &= \frac{u_{\sup,p}\left(1-u_{\sup,p}\right)+\lambda(1+u_{\sup,p}-p)(p-u_{\sup,p})}{\left[f_1\left(F_1^{-1}(u_{\sup,p})\right)\right]^2}, \\
\sigma_U^2 &= \frac{u_{\inf,p}\left(1-u_{\inf,p}\right)+\lambda(u_{\inf,p}-p)(1-u_{\inf,p}+p)}{\left[f_1\left(F_1^{-1}(u_{\inf,p})\right)\right]^2}, \\
\sigma_{LU} &= \frac{\min\left\{u_{\sup,p},u_{\inf,p}\right\}-u_{\sup,p}u_{\inf,p}}{f_1\left(F_1^{-1}(u_{\sup,p})\right)f_1\left(F_1^{-1}(u_{\inf,p})\right)}+ \\
&\quad \lambda\frac{\min\left\{1+u_{\sup,p}-p,u_{\inf,p}-p\right\}-\left(1+u_{\sup,p}-p\right)\left(u_{\inf,p}-p\right)}{f_1\left(F_1^{-1}(u_{\sup,p})\right)f_1\left(F_1^{-1}(u_{\inf,p})\right)}.
\end{aligned}$$

We note that Theorem 5 is valid for a given probability measure P satisfying (A1)-(A4). It is possible to strengthen (A1)-(A4) to convert Theorem 5 to a uniform result in P. We leave this to future work.

Confidence Sets

Confidence Intervals for Each Bound

Given the asymptotic normality of the estimator of each bound, we can apply the standard approach to constructing confidence intervals for each bound. So, the following lemma holds.

Lemma 6 *Let $\hat{f}_1(u)$ be a consistent estimator of $f_1(u)$ and define the followings:*

$$\begin{aligned}
\hat{\sigma}_L^2 &= \frac{\hat{u}_{\sup,p}\left(1-\hat{u}_{\sup,p}\right)+\lambda(1+\hat{u}_{\sup,p}-p)(p-\hat{u}_{\sup,p})}{\left[\hat{f}_1\left(F_{1n}^{-1}(\hat{u}_{\sup,p})\right)\right]^2}, \\
\hat{\sigma}_U^2 &= \frac{\hat{u}_{\inf,p}\left(1-\hat{u}_{\inf,p}\right)+\lambda(\hat{u}_{\inf,p}-p)(1-\hat{u}_{\inf,p}+p)}{\left[\hat{f}_1\left(F_{1n}^{-1}(\hat{u}_{\inf,p})\right)\right]^2}.
\end{aligned}$$

i) Suppose (A1)-(A3) hold. Then,

$$\lim \Pr\left[Q_n^L(p)-z_{1-\alpha/2}\frac{\hat{\sigma}_L}{\sqrt{n_1}} \leq Q^L(p) \leq Q_n^L(p)+z_{1-\alpha/2}\frac{\hat{\sigma}_L}{\sqrt{n_1}}\right]=1-\alpha.$$

ii) Suppose (A1), (A2), and (A4) hold. Then,

$$\lim \Pr\left[Q_n^U(p)-z_{1-\alpha/2}\frac{\hat{\sigma}_U}{\sqrt{n_1}} \leq Q^U(p) \leq Q_n^U(p)+z_{1-\alpha/2}\frac{\hat{\sigma}_U}{\sqrt{n_1}}\right]=1-\alpha.$$

The CIs provided in Lemma 6 for the bounds require the consistent estimation of the density function f_1. Although a kernel density estimator of f_1 may be used, the choice of the bandwidth is troublesome. To avoid the estimation of f_1, we propose a new approach, which extends the well-known confidence intervals for univariate quantiles based on order statistics (see e.g., van der Vaart (1998)) to our case.

There are two problems we have to address in applying that approach for our case. First, the sharp bounds are given by the differences between two univariate quantiles, see (IV.3) and (IV.4). Second, the quantile levels involved are unknown, as $u_{\sup,p}$ and $u_{\inf,p}$ are unknown. We now explain how we handle both problems. The following lemma extends CIs based on order statistics for univariate quantiles (see e.g., Example 21.8 in van der Vaart (1998)) to $Q^L(p)$ and $Q^U(p)$, but assuming $u_{\sup,p}$ and $u_{\inf,p}$ are known. Let $[k]$ be the largest integer that does not exceed k. Define

$$\begin{aligned} u_{0\,\sup,p} &= \frac{1}{n_0} \arg_j \left\{ F_{0n}^{-1}(1 + u_{\sup,p} - p) = Y_{0(j)} \right\}, \\ u_{0\,\inf,p} &= \frac{1}{n_0} \arg_j \left\{ F_{0n}^{-1}(u_{\inf,p} - p) = Y_{0(j)} \right\}, \end{aligned}$$

where $Y_{0(j)}$ is the $j$order statistic of $\{Y_{0i}\}_{i=1}^{n_1}$.

Lemma 7 *i) Suppose (A1)-(A3) hold. Define*

$$
\begin{aligned}
u_{1nL}^{A} &= u_{\sup,p} - z_{1-\alpha/2}\frac{u_{\sup,p}\left(1-u_{\sup,p}\right)}{\sqrt{n_1}\sqrt{u_{\sup,p}\left(1-u_{\sup,p}\right)+\frac{1}{\lambda}u_{0\sup,p}\left(1-u_{0\sup,p}\right)}},\\
u_{0nL}^{A} &= u_{0\sup,p} + z_{1-\alpha/2}\frac{u_{0\sup,p}\left(1-u_{0\sup,p}\right)}{\sqrt{n_1}\sqrt{u_{\sup,p}\left(1-u_{\sup,p}\right)+\frac{1}{\lambda}u_{0\sup,p}\left(1-u_{0\sup,p}\right)}},\\
u_{1nL}^{B} &= u_{\sup,p} + z_{1-\alpha/2}\frac{u_{\sup,p}\left(1-u_{\sup,p}\right)}{\sqrt{n_1}\sqrt{u_{\sup,p}\left(1-u_{\sup,p}\right)+\frac{1}{\lambda}u_{0\sup,p}\left(1-u_{0\sup,p}\right)}},\\
u_{0nL}^{B} &= u_{0\sup,p} - z_{1-\alpha/2}\frac{u_{0\sup,p}\left(1-u_{0\sup,p}\right)}{\sqrt{n_1}\sqrt{u_{\sup,p}\left(1-u_{\sup,p}\right)+\frac{1}{\lambda}u_{0\sup,p}\left(1-u_{0\sup,p}\right)}}.
\end{aligned}
$$

Then

$$
\lim \Pr\left[Y_{1\left(\left[n_1 u_{1nL}^{A}\right]\right)} - Y_{0\left(\left[n_0 u_{0nL}^{A}\right]\right)} \le Q^{L}(p) \le Y_{1\left(\left[n_1 u_{1nL}^{B}\right]\right)} - Y_{0\left(\left[n_0 u_{0nL}^{B}\right]\right)}\right] = 1-\alpha.
$$

ii) Suppose (A1), (A2), and (A4) hold. Define

$$
\begin{aligned}
u_{1nU}^{A} &= u_{\inf,p} - z_{1-\alpha/2}\frac{u_{\inf,p}\left(1-u_{\inf,p}\right)}{\sqrt{n_1}\sqrt{u_{\inf,p}\left(1-u_{\inf,p}\right)+\frac{1}{\lambda}u_{0\inf,p}\left(1-u_{0\inf,p}\right)}},\\
u_{0nU}^{A} &= u_{0\inf,p} + z_{1-\alpha/2}\frac{u_{0\inf,p}\left(1-u_{0\inf,p}\right)}{\sqrt{n_1}\sqrt{u_{\inf,p}\left(1-u_{\inf,p}\right)+\frac{1}{\lambda}u_{0\inf,p}\left(1-u_{0\inf,p}\right)}},\\
u_{1nU}^{B} &= u_{\inf,p} + z_{1-\alpha/2}\frac{u_{\inf,p}\left(1-u_{\inf,p}\right)}{\sqrt{n_1}\sqrt{u_{\inf,p}\left(1-u_{\inf,p}\right)+\frac{1}{\lambda}u_{0\inf,p}\left(1-u_{0\inf,p}\right)}},\\
u_{0nU}^{B} &= u_{0\inf,p} - z_{1-\alpha/2}\frac{u_{0\inf,p}\left(1-u_{0\inf,p}\right)}{\sqrt{n_1}\sqrt{u_{\inf,p}\left(1-u_{\inf,p}\right)+\frac{1}{\lambda}u_{0\inf,p}\left(1-u_{0\inf,p}\right)}}.
\end{aligned}
$$

Then

$$
\lim \Pr\left[Y_{1\left(\left[n_1 u_{1nU}^{A}\right]\right)} - Y_{0\left(\left[n_0 u_{0nU}^{A}\right]\right)} \le Q^{U}(p) \le Y_{1\left(\left[n_1 u_{1nU}^{B}\right]\right)} - Y_{0\left(\left[n_0 u_{0nU}^{B}\right]\right)}\right] = 1-\alpha.
$$

We note that in general it's very hard to extend the confidence intervals for univariate quantiles based on order statistics to other settings including the difference between two quantiles. In our case, $Q^L(p)$ and $Q^U(p)$ are defined by differences between two quantiles, but the quantile levels in the respective quantiles are related by the first order conditions. For example, consider $Q^L(p) = F_1^{-1}\left(u_{\sup,p}\right) - F_0^{-1}\left(1 + u_{\sup,p} - p\right)$. The two quantile levels

$u_{\sup,p}$ and $[1+u_{\sup,p}-p]$ satisfy the following first order condition:

$$f_1\left(F_1^{-1}\left(u_{\sup,p}\right)\right) = f_0\left(F_0^{-1}\left(1+u_{\sup,p}-p\right)\right),$$

which is why we are able to construct confidence intervals in Lemma 7.

In the next result, we show that estimating $u_{\sup,p}$ and $u_{\inf,p}$ in the confidence intervals in Lemma 7 by $\hat{u}_{\sup,p}$ and $\hat{u}_{\inf,p}$ does not affect their validity. Define

$$\begin{aligned}
\hat{u}_{0\sup,p} &= \frac{1}{n_0}\arg_j\left\{F_{0n}^{-1}(1+\hat{u}_{\sup,p}-p)=Y_{0(j)}\right\}, \\
\hat{u}_{0\inf,p} &= \frac{1}{n_0}\arg_j\left\{F_{0n}^{-1}(\hat{u}_{\inf,p}-p)=Y_{0(j)}\right\},
\end{aligned}$$

Theorem 6 *i) Suppose (A1)-(A3) hold. Define*

$$\begin{aligned}
\hat{u}_{1nL}^{A} &= \hat{u}_{\sup,p} - z_{1-\alpha/2}\frac{\hat{u}_{\sup,p}\left(1-\hat{u}_{\sup,p}\right)}{\sqrt{n_1}\sqrt{\hat{u}_{\sup,p}\left(1-\hat{u}_{\sup,p}\right)+\frac{1}{\lambda}\hat{u}_{0\sup,p}\left(1-\hat{u}_{0\sup,p}\right)}}, \\
\hat{u}_{0nL}^{A} &= \hat{u}_{0\sup,p} + z_{1-\alpha/2}\frac{\hat{u}_{0\sup,p}\left(1-\hat{u}_{0\sup,p}\right)}{\sqrt{n_1}\sqrt{\hat{u}_{\sup,p}\left(1-\hat{u}_{\sup,p}\right)+\frac{1}{\lambda}\hat{u}_{0\sup,p}\left(1-\hat{u}_{0\sup,p}\right)}}, \\
\hat{u}_{1nL}^{B} &= \hat{u}_{\sup,p} + z_{1-\alpha/2}\frac{\hat{u}_{\sup,p}\left(1-\hat{u}_{\sup,p}\right)}{\sqrt{n_1}\sqrt{\hat{u}_{\sup,p}\left(1-\hat{u}_{\sup,p}\right)+\frac{1}{\lambda}\hat{u}_{0\sup,p}\left(1-\hat{u}_{0\sup,p}\right)}}, \\
\hat{u}_{0nL}^{B} &= \hat{u}_{0\sup,p} - z_{1-\alpha/2}\frac{\hat{u}_{0\sup,p}\left(1-\hat{u}_{0\sup,p}\right)}{\sqrt{n_1}\sqrt{\hat{u}_{\sup,p}\left(1-\hat{u}_{\sup,p}\right)+\frac{1}{\lambda}\hat{u}_{0\sup,p}\left(1-\hat{u}_{0\sup,p}\right)}}.
\end{aligned}$$

Then

$$\lim \Pr\left[Y_{1([n_1\hat{u}_{1nL}^{A}])} - Y_{0([n_0\hat{u}_{0nL}^{A}])} \le Q^L(p) \le Y_{1([n_1\hat{u}_{1nL}^{B}])} - Y_{0([n_0\hat{u}_{0nL}^{B}])}\right] = 1-\alpha.$$

ii) Suppose (A1), (A2), and (A4) hold. Define

$$\begin{aligned}
\hat{u}_{1nU}^{A} &= \hat{u}_{\inf,p} - z_{1-\alpha/2}\frac{\hat{u}_{\inf,p}\left(1-\hat{u}_{\inf,p}\right)}{\sqrt{n_1}\sqrt{\hat{u}_{\inf,p}\left(1-\hat{u}_{\inf,p}\right)+\frac{1}{\lambda}\hat{u}_{0\inf,p}\left(1-\hat{u}_{0\inf,p}\right)}}, \\
\hat{u}_{0nU}^{A} &= \hat{u}_{0\inf,p} + z_{1-\alpha/2}\frac{\hat{u}_{0\inf,p}\left(1-\hat{u}_{0\inf,p}\right)}{\sqrt{n_1}\sqrt{\hat{u}_{\inf,p}\left(1-\hat{u}_{\inf,p}\right)+\frac{1}{\lambda}\hat{u}_{0\inf,p}\left(1-\hat{u}_{0\inf,p}\right)}}, \\
\hat{u}_{1nU}^{B} &= \hat{u}_{\inf,p} + z_{1-\alpha/2}\frac{\hat{u}_{\inf,p}\left(1-\hat{u}_{\inf,p}\right)}{\sqrt{n_1}\sqrt{\hat{u}_{\inf,p}\left(1-\hat{u}_{\inf,p}\right)+\frac{1}{\lambda}\hat{u}_{0\inf,p}\left(1-\hat{u}_{0\inf,p}\right)}}, \\
\hat{u}_{0nU}^{B} &= \hat{u}_{0\inf,p} - z_{1-\alpha/2}\frac{\hat{u}_{0\inf,p}\left(1-\hat{u}_{0\inf,p}\right)}{\sqrt{n_1}\sqrt{\hat{u}_{\inf,p}\left(1-\hat{u}_{\inf,p}\right)+\frac{1}{\lambda}\hat{u}_{0\inf,p}\left(1-\hat{u}_{0\inf,p}\right)}}.
\end{aligned}$$

Then

$$\lim \Pr\left[Y_{1\left(\left[n_1 \hat{u}^A_{1nU}\right]\right)} - Y_{0\left(\left[n_0 \hat{u}^A_{0nU}\right]\right)} \leq Q^U(p) \leq Y_{1\left(\left[n_1 \hat{u}^B_{1nU}\right]\right)} - Y_{0\left(\left[n_0 \hat{u}^B_{0nU}\right]\right)}\right] = 1 - \alpha.$$

In contrast to the CIs in Lemma 6, the CIs in Theorem 7 do not require estimating f_1.

Confidence Sets for $Q_{\mathbf{TE}}(p)$

Lemma 5 shows that for a given quantile level p, the true quantile $Q_{\mathrm{TE}}(p)$ is interval identified. Various inference procedures have been developed lately for partially identified parameters including interval identified parameters as a special case. We'll apply one of the CSs developed in Chapter II to $Q_{\mathrm{TE}}(p)$ to show that direct applications of the existing CSs require consistent estimation of the density function f_1. To avoid this, we combine the idea of Chernozhukov, Lee, and Rosen (2007) and that used in subsection 3.1 to construct a new CS for $Q_{\mathrm{TE}}(p)$.

Fan and Park's Approach

For notational simplicity, we let $\theta = Q_{\mathrm{TE}}(p)$ and Θ the identification region for θ, i.e., $\Theta = \left[Q^L(p), Q^U(p)\right]$. Define

$$T_n(\theta) = n_1 \left(\frac{Q_n^L(p) - \theta}{\hat{\sigma}_L}\right)_+^2 + n_1 \left(\frac{Q_n^U(p) - \theta}{\hat{\sigma}_U}\right)_-^2$$

and

$$J_{(h_l, h_u, \rho)}(x) = \Pr\left[(Z_{l,\rho} - h_l)_+^2 + (Z_{u,\rho} + h_u)_-^2 \leq x\right]$$

where $x_+ = \max\{x, 0\}$ and $x_- = \min\{x, 0\}$ and $(Z_{l,\rho}, Z_{u,\rho})'$ follows the standard bivariate normal distribution with correlation coefficient ρ.

In Chapter II, we proposed several CSs for interval identified parameters that are asymptotically uniformly valid. One of these CSs applied to our context is:

$$CI_{\mathrm{FP}}=\left\{\theta : T_n(\theta)\leq \max\left\{J^{-1}_{\left(0,\frac{\sqrt{n}\Delta^*}{\hat{\sigma}_U},\hat{\rho}\right)}(1-\alpha)\,,J^{-1}_{\left(\frac{\sqrt{n}\Delta^*}{\hat{\sigma}_L},0,\hat{\rho}\right)}(1-\alpha)\right\}\right\}, \tag{IV.5}$$

where $\hat{\rho}=\frac{\hat{\sigma}_{LU}}{\hat{\sigma}_L\hat{\sigma}_U}$,

$$\begin{aligned}\hat{\sigma}_{LU} &= \frac{\min\left\{\hat{u}_{\sup,p},\hat{u}_{\inf,p}\right\}-\hat{u}_{\sup,p}\hat{u}_{\inf,p}}{\hat{f}_1\left(F_{1n}^{-1}(\hat{u}_{\sup,p})\right)\hat{f}_1\left(F_{1n}^{-1}(\hat{u}_{\inf,p})\right)}\\ &\quad +\lambda\frac{\min\left\{1+\hat{u}_{\sup,p}-p,\hat{u}_{\inf,p}-p\right\}-\left(1+\hat{u}_{\sup,p}-p\right)\left(\hat{u}_{\inf,p}-p\right)}{\hat{f}_1\left(F_{1n}^{-1}(\hat{u}_{\sup,p})\right)\hat{f}_1\left(F_{1n}^{-1}(\hat{u}_{\inf,p})\right)},\\ \Delta^* &= \begin{cases} Q_n^U(p)-Q_n^L(p) \text{ if } Q_n^U(p)-Q_n^L(p)>b_n,\\ 0 \text{ if otherwise,}\end{cases}\end{aligned}$$

for b_n that satisfies $b_n\to 0$ and $\sqrt{n}b_n\to\infty$.

Extension of 'New Approach'

The confidence sets for the true quantile $Q_{\mathrm{TE}}(p)$ presented in the previous subsection depend on a consistent estimation of the density function f_1. To avoid it, we combine the idea in Chernozhukov, Lee, and Rosen (2007) with that in Theorem 7.

Define

$$\begin{aligned}\gamma_n &= \Phi\left(-d_n 1_{\{Q_n^U(p)-Q_n^L(p)>0\}}\left(Q_n^U(p)-Q_n^L(p)\right)\right),\\ \tilde{\alpha}_n &= (1-\gamma_n)\,\alpha\end{aligned}$$

for a d_n such that $d_n\to\infty$ and $d_n/\sqrt{n_1}\to 0$. Chernozhukov, Lee, and Rosen (2007) showed that

$$\lim_{n_1\to\infty}\inf_{\theta\in\Theta}\Pr\left\{\theta\in\left[\hat{k}^l_{1-\tilde{\alpha}_n},\hat{k}^u_{1-\tilde{\alpha}_n}\right]\right\}=1-\alpha,$$

where

$$\begin{aligned}
\hat{k}^{l}_{1-\tilde{\alpha}_n} &= Q^{L}_{n}(p) - \Phi^{-1}(1-\tilde{\alpha}_n)\frac{\hat{\sigma}_L}{\sqrt{n_1}}, \\
\hat{k}^{u}_{1-\tilde{\alpha}_n} &= Q^{U}_{n}(p) + \Phi^{-1}(1-\tilde{\alpha}_n)\frac{\hat{\sigma}_U}{\sqrt{n_1}}.
\end{aligned} \tag{IV.6}$$

One fundamental difference between (IV.5) and (IV.6) is that (IV.5) depends on a consistent estimator of the correlation coefficient ρ, while (IV.6) does not. Let

$$\begin{aligned}
\hat{u}^{A}_{1nL} &= \hat{u}_{\sup,p} - \Phi^{-1}(1-\tilde{\alpha}_n)\frac{\hat{u}_{\sup,p}(1-\hat{u}_{\sup,p})}{\sqrt{n_1}\sqrt{\hat{u}_{\sup,p}(1-\hat{u}_{\sup,p}) + \frac{1}{\lambda}\hat{u}_{0\sup,p}(1-\hat{u}_{0\sup,p})}}, \\
\hat{u}^{A}_{0nL} &= \hat{u}_{0\sup,p} + \Phi^{-1}(1-\tilde{\alpha}_n)\frac{\hat{u}_{0\sup,p}(1-\hat{u}_{0\sup,p})}{\sqrt{n_1}\sqrt{\hat{u}_{\sup,p}(1-\hat{u}_{\sup,p}) + \frac{1}{\lambda}\hat{u}_{0\sup,p}(1-\hat{u}_{0\sup,p})}},
\end{aligned}$$

$$\begin{aligned}
\hat{u}^{B}_{1nU} &= \hat{u}_{\inf,p} + \Phi^{-1}(1-\tilde{\alpha}_n)\frac{\hat{u}_{\inf,p}(1-\hat{u}_{\inf,p})}{\sqrt{n_1}\sqrt{\hat{u}_{\inf,p}(1-\hat{u}_{\inf,p}) + \frac{1}{\lambda}\hat{u}_{0\inf,p}(1-\hat{u}_{0\inf,p})}}, \\
\hat{u}^{B}_{0nU} &= \hat{u}_{0\inf,p} - \Phi^{-1}(1-\tilde{\alpha}_n)\frac{\hat{u}_{0\inf,p}(1-\hat{u}_{0\inf,p})}{\sqrt{n_1}\sqrt{\hat{u}_{\inf,p}(1-\hat{u}_{\inf,p}) + \frac{1}{\lambda}\hat{u}_{0\inf,p}(1-\hat{u}_{0\inf,p})}}.
\end{aligned}$$

It follows from the proof of Theorem 6 that

$$\lim \Pr\left[Q^{L}_{n}(p) - \Phi^{-1}(1-\tilde{\alpha}_n)\frac{\hat{\sigma}_L}{\sqrt{n_1}} \le Q^{L}(p)\right] = \lim \Pr\left[Y_{1([n_1\hat{u}^{A}_{1nL}])} - Y_{0([n_0\hat{u}^{A}_{0nL}])} \le Q^{L}(p)\right],$$

$$\lim \Pr\left[Q^{U}(p) \le Q^{U}_{n}(p) + \Phi^{-1}(1-\tilde{\alpha}_n)\frac{\hat{\sigma}_U}{\sqrt{n_1}}\right] = \lim \Pr\left[Q^{U}(p) \le Y_{1([n_1\hat{u}^{B}_{1nL}])} - Y_{0([n_0\hat{u}^{B}_{0nL}])}\right].$$

Consequently, we obtain

$$\lim_{n_1\to\infty}\inf_{Q_{\mathrm{TE}}(p)\in\Theta}\Pr\left\{Q_{\mathrm{TE}}(p) \in \left[Y_{1([n_1\hat{u}^{A}_{1nL}])} - Y_{0([n_0\hat{u}^{A}_{0nL}])}, Y_{1([n_1\hat{u}^{B}_{1nL}])} - Y_{0([n_0\hat{u}^{B}_{0nL}])}\right]\right\} = 1-\alpha.$$

Simulation Study

This section presents results from an extensive simulation study on the finite sample performances of the CIs for the bounds and for the true quantile developed in the previous sections. For each bound, we applied the CI based on the standard asymptotics, see Lemma 6 and the CI based on order statistics in Theorem 6. For the true quantile, we applied CI_{FP} and the new CI based on order statistics. We report their coverage rates and their widths.

To see the effects of the underlying marginal distributions on the performances of these CIs, we used 6 different distributions to generate Y_1 and Y_0. Below are their cdfs and pdfs. Their graphs are provided in Appendix B. <Model 1> is the combination of two normals, <Model 2> is taken from Chapter III, and <Model 3> to <Model 6> are modifications of normal mixtures in Marron and Wand (1992).

<Model 1> $Y_1 \sim N(2,2)$ and $Y_0 \sim N(1,1)$

<Model 2> $Y_1 \sim C\left(\frac{1}{4}\right)$ and $Y_0 \sim C\left(\frac{3}{4}\right)$, where

$$X \sim C(a) \implies P(X \le x) = \begin{cases} \dfrac{1}{a}x^2 \text{ if } x \in [0,a] \\ 1 - \dfrac{(x-1)^2}{(1-a)} \text{ if } x \in [a,1] \end{cases}$$

<Model 3> Two skewed unimodal distributions:

$$\begin{aligned} F_1(x) &= \frac{1}{4}\sum_{i=0}^{3} N\left(3\left(1-\left(\frac{2}{3}\right)^i\right)+\frac{29}{36}, \left(\frac{2}{3}\right)^{2i}\right), \\ F_0(x) &= \frac{1}{2}\sum_{i=0}^{1} N\left(3\left(\left(\frac{2}{3}\right)^i-1\right)+\frac{3}{2}, \left(\frac{2}{3}\right)^{2i}\right). \end{aligned}$$

<Model 4> Two bimodal distributions:

$$\begin{aligned} F_1(x) &= \frac{3}{4}N\left(\frac{5}{4}, \left(\frac{3}{4}\right)^2\right)+\frac{1}{4}N\left(\frac{17}{4}, \left(\frac{1}{4}\right)^2\right), \\ F_0(x) &= \frac{1}{2}N\left(\frac{1}{2}, \left(\frac{1}{3}\right)^2\right)+\frac{1}{2}N\left(\frac{3}{2}, \left(\frac{1}{3}\right)^2\right). \end{aligned}$$

<Model 5> Two kurtotic distributions:

$$\begin{aligned} F_1(x) &= \frac{1}{10}N(2,1) + \frac{9}{10}N\left(2, \left(\frac{1}{10}\right)^2\right), \\ F_0(x) &= \frac{1}{2}N(1,1) + \frac{1}{2}N\left(1, \left(\frac{1}{10}\right)^2\right). \end{aligned}$$

<Model 6> Y_0 follows a skewed unimodal distribution and Y_1 a bimodal distribution:

$$\begin{aligned} F_1(x) &= \frac{3}{4}N\left(\frac{5}{4}, \left(\frac{3}{4}\right)^2\right) + \frac{1}{4}N\left(\frac{17}{4}, \left(\frac{1}{4}\right)^2\right), \\ F_0(x) &= \frac{1}{4}\sum_{i=0}^{3} N\left(3\left(1-\left(\frac{2}{3}\right)^i\right) - \frac{7}{36}, \left(\frac{2}{3}\right)^{2i}\right). \end{aligned}$$

In simulations, we drew $n_1 = n_0 = n = 1000, 2000, 4000,$ and/or 6000 samples of Y_1 and Y_0 and estimated $Q_n^L(p)$ and $Q_n^U(p)$ for $p = 0.1, 0.3, 0.5, 0.7,$ and 0.9. The number of replications is 1,000 for each setting. The nominal coverage level is 95%.

Confidence Intervals for Each Bound

The confidence intervals were constructed for each bound based on the 'new approach' and the standard asymptotic approach. For the latter, we need to estimate $f_1(\cdot)$ and $f_0(\cdot)$. We used the kernel density estimation with three different bandwidths to see the sensitivity of the CIs to the value of the bandwidth. The first bandwidth we used is $h_S = (4/3)^{1/5}\hat{\sigma}n^{-1/5}$ where $\hat{\sigma}$ is the sample standard deviation of Y_1 and/or Y_0. The second is $h_R = (4/3)^{1/5}(R/1.34)\,n^{-1/5}$ where R is the interquartile range of Y_1 and/or Y_0. Lastly, we used $h_M = 0.9Mn^{-1/5}$ where $M = \min\{\hat{\sigma}, R/1.34\}$. The second and the third bandwidths are supposed to work better than h_S if the underlying distributions are long-tailed or skewed[2].

[2] Silverman (1986) pp. 45-48.

To save space, only parts of results and summaries are provided here. Appendix C includes the details. In the tables in Appendix C and below, New stands for the 'new approach' and $A(\cdot)$ the standard asymptotic approach with specified bandwidths in the parentheses. C$(\cdot)$ means the coverage rates and W$(\cdot)$ the width of confidence intervals. For example, the table below shows, for $p = 0.1$ and $n = 1000$ for <Model 1>, the average coverage rates of the confidence intervals of the 'new approach' are 0.943 for $Q^L(p)$ and 0.965 for $Q^U(p)$. The average widths for the confidence intervals are 0.4586 for $Q^L(p)$ and 0.3211 for $Q^U(p)$.

<Model 1>	$p = 0.1$, $n = 1000$			
	C(Q^L)	C(Q^U)	W(Q^L)	W(Q^L)
New	.943	.965	.4586	.3211
$A(h_S)$	.942	.942	.4669	.3248
$A(h_R)$	.942	.942	.4671	.3251
$A(h_M)$	.937	.939	.4620	.3219

In the above table, the methods that gave the closest average coverage rate to the nominal level is New for Q^L, $A(h_S)$ and $A(h_R)$ for Q^U. The method that made the width of confidence interval the smallest is New for both $Q^L(p)$ and $Q^U(p)$. In the following tables, we provide these information in one cell. Each cell contains two brackets. The upper bracket shows the methods that gave the closest coverage rates for $Q^L(p)$ to the left and $Q^U(p)$ to the right. The lower bracket shows the methods that made the narrowest confidence intervals for $Q^L(p)$ to the left and $Q^U(p)$ to the right of the bracket. For example, the cell for <Model 1> for $p = 0.1$ and $n = 1000$ is

$[New, A(h_S, h_R)]$ $[New, New]$

where $A(h_S, h_R)$ is a shorthand notation for $A(h_S)$ and $A(h_R)$. It means the asymptotic approach with h_S and h_R have the same number of cases which have minimum deviation from the nominal level for Q^U.

Table 12. Summary of Model 1

	$p = 0.1$	$p = 0.3$	$p = 0.5$
$n = 1,000$	$[New, A(h_S, h_R)]$ $[New, New]$	$[New, New]$ $[New, New]$	$[A(h_S, h_R), A(h_S, h_R)]$ $[New, New]$
$n = 2,000$	$[New, New]$ $[New, New]$	$[All, A(h_S, h_R)]$ $[New, New]$	$[New, New]$ $[New, New]$
$n = 4,000$	$[All, New]$ $[New, New]$	$[New, New]$ $[New, New]$	$[A(h_S, h_R), New]$ $[New, New]$
$n = 6,000$	$[A(h_S, h_R), New]$ $[New, New]$	$[New, A(h_S, h_R)]$ $[New, New]$	$[New, All]$ $[New, New]$

	$p = 0.7$	$p = 0.9$
$n = 1,000$	$[New, New]$ $[New, New]$	$[A(h_S, h_R), A(h_S)]$ $[New, A(h_M)]$
$n = 2,000$	$[New, New]$ $[New, New]$	$[New, A(h_S, h_R)]$ $[New, A(h_M)]$
$n = 4,000$	$[A(h_S, h_R), A(h_M)]$ $[New, New]$	$[New, A(h_S, h_R)]$ $[New, New]$
$n = 6,000$	$[New, A(h_S, h_R)]$ $[New, New]$	$[New, A(h_S, h_R)]$ $[New, New]$

	C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)
New	14	9	20	18
$A(h_S)$	7	10	0	0
$A(h_R)$	7	9	0	0
$A(h_M)$	2	2	0	2

Table 12 presents the results for <Model 1>. Here, *All* stands 'all methods'. It means all methods gave the same number of best coverage rates or the smallest widths. New appears 14 times in the coverage for $Q^L(p)$, which is twice as many as $A(h_S)$ and $A(h_M)$ do and seven times as $A(h_M)$ does. This implies that the 'new approach' generated the closest coverage rates to the nominal level two times or seven times frequently on average compared to others. As to the coverage rates for Q^U, New, $A(h_S)$, and $A(h_R)$ appear almost the same times (9-10 times). Only $A(h_M)$ is worse than the others. When the widths of confidence intervals are considered, New is outstanding. It produced the minimum widths in 20 (for the width for Q^L) and 18 (for the width for Q^U) out of 20 cases. The reason that New produces the narrowest intervals on average yet covers true values of Q^L and Q^U most accurately is because the confidence intervals constructed from New are asymmetric around Q_n^L or Q_n^U. In all simulation models, the confidence intervals for Q^L have a wider range to the left of Q_n^L than to the right and that for Q^U are the opposite. In summary, (i) New performs the best in $\mathrm{C}(Q^L)$ with the narrowest confidence intervals. (ii) the asymptotic approach denoted by $A(.)$ is sensitive to the value of the bandwidth.

Table 13. Summary of Model 2

	$p = 0.1$	$p = 0.3$	$p = 0.5$	$p = 0.7$	$p = 0.9$
$n = 1,000$	$[New, (New, A(h_R))]$ $[New, A(h_M)]$	$[A(h_M), New]$ $[New, A(h_M)]$	$[A(h_S), New]$ $[New, A(h_M)]$	$[New, New]$ $[New, A(h_M)]$	$[New, A(h_R)]$ $[New, New]$
$n = 2,000$	$[New, New]$ $[New, A(h_M)]$	$[New, New]$ $[New, A(h_M)]$	$[A(h_M), New]$ $[New, A(h_M)]$	$[New, New]$ $[New, A(h_M)]$	$[New, A(h_R)]$ $[New, New]$
$n = 4,000$	$[A(h_M), A(h_S)]$ $[New, A(h_M)]$	$[A(h_R), New]$ $[New, A(h_M)]$	$[A(h_R), All]$ $[New, A(h_M)]$	$[New, A(h_M)]$ $[A(h_M), A(h_M)]$	$[New, New]$ $[New, New]$
$n = 6,000$	$[A(M), New]$ $[New, A(h_M)]$	$[New, All]$ $[New, A(h_M)]$	$[New, New]$ $[New, A(h_M)]$	$[New, New]$ $[New, A(h_M)]$	$[New, A(h_R)]$ $[New, New]$

	$\mathrm{C}(Q^L)$	$\mathrm{C}(Q^U)$	$\mathrm{W}(Q^L)$	$\mathrm{W}(Q^U)$
New	13	13	19	4
$A(h_S)$	1	3	0	0
$A(h_R)$	2	6	0	0
$A(h_M)$	4	3	1	16

Table 13 summarizes the results for <Model 2>. Roughly speaking, the same pattern with <Model 1> exists except for $\mathrm{W}(Q^U)$. New appears most frequently in $\mathrm{C}(Q^L)$, $\mathrm{C}(Q^U)$, and $\mathrm{W}(Q^L)$. In $\mathrm{W}(Q^U)$, $A(h_M)$ is the most frequent. New covers Q^L most accurately with the smallest confidence interval.

Table 14 shows that for <Model 3>. Differently from <Model 1> and <Model 2>, New and $A(h_M)$ are the two dominant methods for different parameters in $\mathrm{C}(Q^L)$ and $\mathrm{C}(Q^U)$ respectively. In $\mathrm{C}(Q^L)$, New, $A(h_S)$, and $A(h_R)$ show similar performance, although New is slightly better. In $\mathrm{C}(Q^U)$, $A(h_M)$ and New are better compared to the other two. In this setting, none of the methods were 'best' in the sense that no methods considered generated most coverage with smallest confidence intervals. For example, $A(h_M)$ that generated the smallest confidence intervals for Q^L most frequently were the worst in $\mathrm{C}(Q^L)$ and $A(h_M)$, the best in $\mathrm{C}(Q^U)$, was 0 in $\mathrm{W}(Q^U)$.

Table 14. Summary of Model 3

	$p = 0.1$	$p = 0.3$	$p = 0.5$	$p = 0.7$	$p = 0.9$
$n = 1,000$	$[New, A(h_M)]$ $[A(h_M), New]$	$[All, A(h_S)]$ $[A(h_M), New]$	$[New, New]$ $[A(h_M), New]$	$[New, A(h_M)]$ $[A(h_M), New]$	$[A(h_S, h_R), (New, A(h_M))]$ $[A(h_M), New]$
$n = 2,000$	$[New, A(h_M)]$ $[A(h_M), New]$	$[A(h_S, h_R, h_M), A(h_R)]$ $[A(h_M), New]$	$[A(h_S, h_R), New]$ $[A(h_M), New]$	$[A(h_M), A(h_M)]$ $[A(h_M), New]$	$[New, A(h_M)]$ $[New, New]$
$n = 4,000$	$[A(h_S, h_R), A(h_M)]$ $[A(h_M), New]$	$[New, A(h_M)]$ $[A(h_M), New]$	$[A(h_S, h_R), A(h_M)]$ $[A(h_M), New]$	$[A(h_S, h_R), (New, A(h_M))]$ $[A(h_M), New]$	$[A(h_S, h_R, h_M), A(h_M)]$ $[A(h_M), New]$
$n = 6,000$	$[A(h_S, h_R, h_M), (New, A(h_R))]$ $[A(h_M), New]$	$[New, A(h_S, h_R)]$ $[A(h_M), New]$	$[New, New]$ $[A(h_M), New]$	$[New, New]$ $[A(h_M), New]$	$[New, New]$ $[A(h_M), New]$

	$\mathrm{C}(Q^L)$	$\mathrm{C}(Q^U)$	$\mathrm{W}(Q^L)$	$\mathrm{W}(Q^U)$
New	11	8	1	20
$A(h_S)$	9	2	0	0
$A(h_R)$	9	3	0	0
$A(h_M)$	5	11	19	0

As presented in Table 15, with <Model 4>, New generated the narrowest confidence intervals for both Q^L and Q^U on average and appears to perform the best in $\mathrm{C}(Q^L)$ and $\mathrm{C}(Q^U)$. In $\mathrm{C}(Q^U)$, $A(h_M)$ is as good as New however other h's were not as good as h_M.

Table 15. Summary of Model 4

	$p = 0.1$	$p = 0.3$	$p = 0.5$
$n = 1{,}000$	$[New, A(h_M)]$ $[New, New]$	$[New, New]$ $[New, New]$	$[New, New]$ $[New, New]$
$n = 2{,}000$	$[New, New]$ $[New, New]$	$[New, A(h_M)]$ $[New, New]$	$[New, (New, A(h_M))]$ $[New, New]$
$n = 4{,}000$	$[New, (New, A(h_M))]$ $[New, New]$	$[A(h_M), New]$ $[New, New]$	$[New, New]$ $[New, New]$
$n = 6{,}000$	$[(New, A(h_M)), A(h_M)]$ $[New, New]$	$[A(h_M), A(h_M)]$ $[New, New]$	$[A(h_S), New]$ $[New, New]$

	$p = 0.7$	$p = 0.9$
$n = 1{,}000$	$[New, New]$ $[New, A(h_R)]$	$[New, A(h_S, h_R, h_M)]$ $[New, New]$
$n = 2{,}000$	$[New, A(h_S)]$ $[New, A(h_R)]$	$[New, A(h_S, h_R, h_M)]$ $[New, New]$
$n = 4{,}000$	$[New, New]$ $[New, New]$	$[New, A(h_S, h_R, h_M)]$ $[New, New]$
$n = 6{,}000$	$[A(h_M), New]$ $[New, New]$	$[New, A(h_S, h_R, h_M)]$ $[New, New]$

	$\mathrm{C}(Q^L)$	$\mathrm{C}(Q^U)$	$\mathrm{W}(Q^L)$	$\mathrm{W}(Q^U)$
New	16	11	20	18
$A(h_S)$	1	5	0	0
$A(h_R)$	0	4	0	2
$A(h_M)$	4	10	0	0

For <Model 5>, New is better than others (in $\mathrm{C}(Q^L)$) or as good as others (in $\mathrm{C}(Q^U)$, $\mathrm{W}(Q^L)$, and/or $\mathrm{W}(Q^U)$). $A(h_S)$ works poorly in $\mathrm{C}(Q^L)$ and $\mathrm{C}(Q^U)$ while $A(h_M)$ works as good as New in $\mathrm{C}(Q^U)$. See Table 16.

Table 16. Summary of Model 5

	$p=0.1$	$p=0.3$	$p=0.5$
$n=1,000$	$[New, New]$ $[A(h_S), A(h_S)]$	$[A(h_R), New]$ $[A(h_M), New]$	$[New, New]$ $[New, New]$
$n=2,000$	$[New, A(h_M)]$ $[A(h_S), A(h_S)]$	$[(New, A(h_R)), New]$ $[A(h_M), New]$	$[New, A(h_M)]$ $[New, New]$
$n=4,000$	$[New, A(h_S)]$ $[A(h_S), A(h_S)]$	$[A(h_R), A(h_M)]$ $[A(h_M), New]$	$[New, A(h_M)]$ $[New, New]$
$n=6,000$	$[New, A(h_R)]$ $[A(h_S), A(h_S)]$	$[A(h_S), (New, A(h_R))]$ $[A(h_M), New]$	$[A(h_R), A(h_R)]$ $[New, New]$

	$p=0.7$	$p=0.9$
$n=1,000$	$[A(h_M), New]$ $[New, A(h_M)]$	$[New, A(h_R)]$ $[A(h_S), A(h_S)]$
$n=2,000$	$[New, A(h_R, h_M)]$ $[New, A(h_M)]$	$[New, New]$ $[A(h_S), A(h_S)]$
$n=4,000$	$[A(h_M), A(h_M)]$ $[New, A(h_M)]$	$[A(h_R, h_M), New]$ $[A(h_S), A(h_S)]$
$n=6,000$	$[(New, A(h_M)), A(h_R, h_M)]$ $[New, A(h_M)]$	$[New, A(h_M)]$ $[A(h_S), A(h_S)]$

	$\mathrm{C}(Q^L)$	$\mathrm{C}(Q^U)$	$\mathrm{W}(Q^L)$	$\mathrm{W}(Q^U)$
New	13	8	8	8
$A(h_S)$	1	1	8	8
$A(h_R)$	5	6	0	0
$A(h_M)$	4	8	4	4

In <Model 6>, New clearly outperforms the others. See Table 17.

Table 17. Summary of Model 6

	$p=0.1$	$p=0.3$	$p=0.5$	$p=0.7$	$p=0.9$
$n=1,000$	$[New, A(h_M)]$ $[New, New]$	$[New, A(h_M)]$ $[New, New]$	$[New, A(h_M)]$ $[New, New]$	$[New, New]$ $[New, A(h_R)]$	$[New, New]$ $[New, New]$
$n=2,000$	$[New, New]$ $[New, New]$	$[New, A(h_S)]$ $[New, New]$	$[New, NewA(h_M)]$ $[New, New]$	$[New, A(h_R)]$ $[New, A(h_R)]$	$[New, A(h_R)]$ $[New, New]$
$n=4,000$	$[New, A(h_M)]$ $[New, New]$	$[A(h_M), New]$ $[New, New]$	$[A(h_S), New]$ $[New, New]$	$[New, A(h_R)]$ $[New, A(h_M)]$	$[New, New]$ $[New, New]$
$n=6,000$	$[(New, A(h_M)), A(h_S, h_M)]$ $[New, New]$	$[A(h_M), New]$ $[New, New]$	$[A(h_S, h_R), New]$ $[New, New]$	$[A(h_M), A(h_R)]$ $[New, A(h_M)]$	$[New, (New, A(h_M))]$ $[New, New]$

	C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)
New	15	10	20	16
$A(h_S)$	2	2	0	0
$A(h_R)$	1	4	0	2
$A(h_M)$	4	7	0	2

From the simulation results, we can see: i) the 'new approach' covers the Q^L and Q^U most accurately in general; ii) in some cases such as Q^L in <Model 1>, <Model 2>, <Model 4>, <Model 5>, <Model 6> and Q^U in <Model 4>, <Model 5>, and <Model 6>, the 'new approach' was the best in the sense that not only it covers the true values most accurately but also it does that with the smallest confidence intervals; iii) the choice of bandwidth changes the performance of the asymptotic method but there is no generalizable pattern. In addition to the simulation results, it makes the 'new approach' preferable that the 'new approach' doesn't require estimating $f_1(\cdot)$ where as the results based on the standard asymptotics are sensitive to the choice of the bandwidth.

Confidence Intervals for $Q_{\mathbf{TE}}(p)$

This section provides the simulation results on the coverage rates and the widths of confidence intervals for the true quantiles for <Model 1> through <Model 6>. The methods implemented in this subsection are: CI_{FP} and the modified 'new approach' (denoted as New).

In the simulations, we modified the definition of the shrinkage estimator for Δ like in Chapter II. Originally,

$$\Delta^* = \begin{cases} Q_n^U(p) - Q_n^L(p) \text{ if } Q_n^U(p) - Q_n^L(p) > b_n, \\ 0 \text{ if otherwise.} \end{cases}$$

However, noticing that $Q_n^U(p) - Q_n^L(p)$ can be arbitrarily large by choosing different measurement units, we defined

$$\Delta^* = \begin{cases} Q_n^U(p) - Q_n^L(p) \text{ if } Q_n^U(p) - Q_n^L(p) > b_n \hat{\sigma}_\Delta, \\ 0 \text{ if otherwise,} \end{cases}$$

where $\hat{\sigma}_\Delta^2 \equiv \hat{\sigma}_L^2 + \hat{\sigma}_U^2 - 2\hat{\sigma}_{LU}$. Having used $\hat{\sigma}_\Delta^2$, $\hat{\sigma}_L^2$, and $\hat{\sigma}_U^2$, the confidence intervals from CI_{FP} are subject to the choice of the bandwidth h. In the tables in Appendix D, we used $CI_{\mathrm{FP}}(h_s)$, $CI_{\mathrm{FP}}(h_R)$, and/or $CI_{\mathrm{FP}}(h_M)$ to distinguish among different h's. The New needs not be distinguished because New does not depend on h. We used $b_n = n_1^{-1/3}$ and $d_n = 1/b_n$ in the simulations.

The coverage rates are computed for $Q(p) = Q^L(p)$, $Q^U(p)$, and $\left(Q^L(p) + Q^U(p)\right)/2$. We used $\mathrm{C}\left(Q^L\right)$ to indicate the coverage rates for $Q(p) = Q^L(p)$ and $\mathrm{C}\left(Q^U\right)$ for $Q(p) = Q^U(p)$ but, since the coverage rates for $Q(p) = \left(Q^L(p) + Q^U(p)\right)/2$ were always 1, we omitted $\mathrm{C}\left(\left(Q^L(p) + Q^U(p)\right)/2\right)$. We present a summary here and provide detailed tables in Appendix D. Below is a representative table.

<Model 1>	$p = 0.1$, $n = 1,000$		
	$\mathrm{C}(Q^L)$	$\mathrm{C}(Q^U)$	$\mathrm{W}(Q)$
New	.949	.941	4.3596
$CI_{\mathrm{FP}}\ (h_S)$	.901	.884	4.2726
$CI_{\mathrm{FP}}\ (h_R)$	.902	.883	4.2728
$CI_{\mathrm{FP}}\ (h_M)$	.894	.882	4.2693

For $p = 0.1$ and $n = 1,000$ in <Model 1>, the minimum of the average coverage rate of the New method is 0.941, which is the closest of all methods to the nominal level. In the followings, we will say this the best minimum coverage rate to indicate the closest minimum coverage rates to the nominal level. With this expression, it can be said that New presented the best minimum coverage rate for $p = 0.1$ and $n = 1,000$ in <Model 1>. The average width of confidence intervals generated by $CI_{\mathrm{FP}}\ (h_M)$ is the smallest. As in the previous subsection, we summarize this result as $[New, CI_{\mathrm{FP}}\ (h_M)]$, i.e. the left entry in the bracket shows the method by which the best minimum coverage rate was generated and the right entry is the method by which the smallest average width of confidence intervals.

In <Model 1> - <Model 3>, as presented in Table 18, New outperformed others because it provided the best minimum coverage rates in all settings although it did not generated the smallest confidence intervals. The smallest confidence intervals were generated by $CI_{\mathrm{FP}}\ (h_M)$ in all settings.

Table 19 presents the results of <Model 4>. New provided the best minimum coverage rates in 19 settings. Of the four settings that New generated the smallest confidence intervals, it provided the best minimum coverage rates in three settings. $CI_{\mathrm{FP}}\ (h_M)$ generated the smallest confidence intervals on average in 15 settings.

In <Model 5>, as presented in Table 20, the results are slightly complex. New provided the best minimum coverage rates in 17 settings with providing the smallest confidence intervals in none of them. Contrary to <Model 1> - <Model 4>, $CI_{\mathrm{FP}}\ (h_S)$ generated

Table 18. Summary of Model 1 - Model 3

	$p = 0.1$	$p = 0.3$	$p = 0.5$
$n = 1,000$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$
$n = 2,000$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$
$n = 4,000$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$
$n = 6,000$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$

	$p = 0.7$	$p = 0.9$
$n = 1,000$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$
$n = 2,000$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$
$n = 4,000$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$
$n = 6,000$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$

	New	$CI_{\text{FP}}(h_S)$	$CI_{\text{FP}}(h_R)$	$CI_{\text{FP}}(h_M)$
minC(Q)	20	0	0	0
W(Q)	0	0	0	20

Table 19. Summary of Model 4

	$p = 0.1$	$p = 0.3$	$p = 0.5$
$n = 1,000$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$
$n = 2,000$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$
$n = 4,000$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$
$n = 6,000$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$

	$p = 0.7$	$p = 0.9$
$n = 1,000$	$[New, CI_{\text{FP}}(h_R)]$	$[New, New]$
$n = 2,000$	$[New, CI_{\text{FP}}(h_M)]$	$[CI_{\text{FP}}(h_M), New]$
$n = 4,000$	$[New, CI_{\text{FP}}(h_M)]$	$[New, New]$
$n = 6,000$	$[New, CI_{\text{FP}}(h_M)]$	$[New, New]$

	New	$CI_{\text{FP}}(h_S)$	$CI_{\text{FP}}(h_R)$	$CI_{\text{FP}}(h_M)$
minC(Q)	19	0	0	1
W(Q)	4	0	0	15

the best minimum coverage rates in 5 settings, yet none of which are accompanied with the smallest confidence interval. $CI_{\text{FP}}(h_M)$ generated the smallest confidence intervals on average in 12 settings.

The results of <Model 6> are quite similar to that of <Model 1> - <Model 3>. New presented the best minimum coverage rates in 18 settings and $CI_{\text{FP}}(h_M)$ generated the smallest confidence intervals in 19 settings. See Table 21.

To summarize, New outperformed the other methods because it presented the best minimum coverage rates as was the case for each bound. Contrary to each bound,

Table 20. Summary of Model 5

	$p = 0.1$	$p = 0.3$	$p = 0.5$
$n = 1,000$	$[New, CI_{\text{FP}}(h_S)]$	$[CI_{\text{FP}}(h_S), CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$
$n = 2,000$	$[New, CI_{\text{FP}}(h_S)]$	$[(New, CI_{\text{FP}}(h_S)), CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$
$n = 4,000$	$[New, CI_{\text{FP}}(h_S)]$	$[(New, CI_{\text{FP}}(h_S)), CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$
$n = 6,000$	$[New, CI_{\text{FP}}(h_S)]$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$

	$p = 0.7$	$p = 0.9$
$n = 1,000$	$[CI_{\text{FP}}(h_S), CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_S)]$
$n = 2,000$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_S)]$
$n = 4,000$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_S)]$
$n = 6,000$	$[CI_{\text{FP}}(h_S), CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_S)]$

	New	$CI_{\text{FP}}(h_S)$	$CI_{\text{FP}}(h_R)$	$CI_{\text{FP}}(h_M)$
minC(Q)	17	5	0	0
W(Q)	0	8	0	12

Table 21. Summary of Model 6

	$p = 0.1$	$p = 0.3$	$p = 0.5$
$n = 1,000$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$
$n = 2,000$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$
$n = 4,000$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$	$[CI_{\text{FP}}(h_R), CI_{\text{FP}}(h_M)]$
$n = 6,000$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$

	$p = 0.7$	$p = 0.9$
$n = 1,000$	$[New, CI_{\text{FP}}(h_R)]$	$[New, CI_{\text{FP}}(h_M)]$
$n = 2,000$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$
$n = 4,000$	$[New, CI_{\text{FP}}(h_M)]$	$[New, CI_{\text{FP}}(h_M)]$
$n = 6,000$	$[New, CI_{\text{FP}}(h_M)]$	$[CI_{\text{FP}}(h_M), CI_{\text{FP}}(h_M)]$

	New	$CI_{\text{FP}}(h_S)$	$CI_{\text{FP}}(h_R)$	$CI_{\text{FP}}(h_M)$
minC(Q)	18	0	1	1
W(Q)	0	0	1	19

though, the good performance of New in terms of coverage rate did not go along with the performance in terms of confidence interval width. New appeared both in the left and right entry at the same time in only three settings out of 120 in total. Another notable difference to the each bound case is that CI_{FP} were not as sensitive to the choice of h. $CI_{\text{FP}}(h_M)$ generated the smallest confidence interval in general in all models and $CI_{\text{FP}}(h_S)$ and $CI_{\text{FP}}(h_R)$ did not contribute much except for <Model 5> where $CI_{\text{FP}}(h_S)$ achieved the best minimum coverage rates in 5 settings and smallest confidence intervals in 8 settings. The Monte Carlo study in this section provided evidences that the 'new approach' for both each bound and the true quantile is better than other methods considered in terms of performance in coverage rates and in that it is not subject to the choice of the bandwidth.

Conclusion

This chapter is the first to develop nonparametric estimation and inference procedures for sharp bounds on the quantile of the effect of a binary treatment defined as the difference between the two potential outcomes. In addition to CIs based on the standard asymptotics, we construct novel CIs for both the bounds and the true quantile of treatment effect that avoid the estimation of the marginal density functions. Extensive simulation results show that the new CIs outperformed the ones based on the standard asymptotics and the performance of the latter is sensitive to the choice of the bandwidth.

Much work remains to be done. In terms of the sharp bounds, those in this chapter are the worst bounds in the sense that they do not make use of any prior information on the possible dependence between the potential outcomes. When such information is available, these bounds can be tightened. The focus on randomized experiments in this chapter allows the identification of the marginal distributions. In cases where the marginal distributions themselves are not identifiable but bounds on them can be placed (see, e.g., Manski (1994, 2003), Manski and Pepper (2000), Shaikh and Vytlacil (2005), Blundell, Gosling, Ichimura, and Meghir (2006), Honore and Lleras-Muney (2007)), we can also place bounds on the quantile function of the treatment effect.

In terms of statistical inference, the results developed in this chapter rely on the uniqueness of $u_{\sup,p}$ and $u_{\inf,p}$ as assumed in (A3) and (A4). Although the simulation results in this chapter show that the new CIs performed very well for both the bounds themselves and the true quantile, it remains to see if this condition might be relaxed.

Appendix A. Technical Proofs

<Proof for Theorem 5>

We follow the same steps as in the proof of Proposition 1 in Chapter III. We prove the Q^L part only. Let

$$\Psi_n(u,p) = F_{1n}^{-1}(u) - F_{0n}^{-1}(1+u-p)\,,\ \Psi(u,p) = F_1^{-1}(u) - F_0^{-1}(1+u-p)\,.$$

$Q_n^L(p) = \Psi_n(\hat{u}_{\sup,p}, p)$ and $Q^L(p) = \Psi(u_{\sup,p}, p)$. Let $\overline{Q}_n(p) = \Psi_n(u_{\sup,p}, p)$. (A2) guarantees $\sqrt{n_1}\left(\overline{Q}_n(p) - Q^L(p)\right) \Longrightarrow N\left(0, \sigma_L^2\right)$ (see Csrögő 1983, p.10). We will complete the proof of (i) in three steps:

1. We show that $\hat{u}_{\sup,p} - u_{\sup,p} = o_p(1)$;
2. We show that $\hat{u}_{\sup,p} - u_{\sup,p} = O_p\left(n_1^{-1/3}\right)$;
3. $\sqrt{n_1}\left(Q_n^L(p) - Q^L(p)\right)$ has the same limiting distribution as $\sqrt{n_1}\left(\overline{Q}_n(p) - Q^L(p)\right)$.

Proof of 1. By Corollary 1.4.2 in Csrögő (1983),

$$\sup_{\frac{1}{n_1}\le u\le 1-\frac{1}{n_1}}\left|F_{1n}^{-1}(u) - F_1^{-1}(u)\right| \to 0,\ \sup_{\frac{1}{n_0}\le u\le 1-\frac{1}{n_0}}\left|F_{0n}^{-1}(u) - F_0^{-1}(u)\right| \to 0.$$

Therefore,

$$\begin{aligned}
& \sup_{\frac{1}{n_1}\le u\le p}\left|\left[F_{1n}^{-1}(u) - F_{0n}^{-1}(1+u-p)\right] - \left[F_1^{-1}(u) - F_0^{-1}(1+u-p)\right]\right| \\
= & \sup_{\frac{1}{n_1}\le u\le 1-\frac{1}{n_1}}\left|\left[F_{1n}^{-1}(u) - F_1^{-1}(u)\right] - \left[F_{0n}^{-1}(1+u-p) - F_0^{-1}(1+u-p)\right]\right| \\
\le & \sup_{\frac{1}{n_1}\le u\le 1-\frac{1}{n_1}}\left|\left[F_{1n}^{-1}(u) - F_1^{-1}(u)\right]\right| + \sup_{\frac{1}{n_0}\le u\le 1-\frac{1}{n_0}}\left|\left[F_{0n}^{-1}(1+u-p) - F_0^{-1}(1+u-p)\right]\right| \\
\to & \ 0.
\end{aligned}$$

This and A3(i) imply that the sequence $\hat{u}_{\sup,p}$ converges in probability to $u_{\sup,p}$, see e.g., Theorem 5.7 in van der Vaart (1998).

Proof of 2. We use Theorem 3.2.5 in van der Vaart and Wellner (1996) to establish the rate of convergence for $\hat{u}_{\sup,p}$. Given (A2) and (A3), the map: $u \mapsto \Psi(u,p)$ is twice differentiable and has a unique maximum at $u_{\sup,p}$. By (A3), the first condition of Theorem 3.2.5 in van der Vaart and Wellner (1996) is satisfied with $\alpha = 2$. To check the second condition of Theorem 3.2.5 in van der Vaart and Wellner (1996), we consider the centered process:

$$\begin{aligned}
& \sqrt{n_1}\left(\Psi_n - \Psi\right)(u,p) \\
= \;& \sqrt{n_1}\left(F_{1n}^{-1} - F_1^{-1}\right)(u) - \sqrt{n_1}\left(F_{0n}^{-1} - F_0^{-1}\right)(u+1-p) \\
= \;& -\frac{1}{f_1\left(F_1^{-1}(u)\right)}\sqrt{n_1}\left(F_{1n}\left(F_1^{-1}(u)\right) - u\right) \\
& + \frac{1}{f_0\left(F_0^{-1}(1+u-p)\right)}\frac{\sqrt{n_1}}{\sqrt{n_0}}\sqrt{n_0}\left(F_{0n}\left(F_0^{-1}(u+1-p)\right) - (u+1-p)\right) \\
& + o_p(1) \\
\equiv \;& G_{n1}(u) - \frac{\sqrt{n_1}}{\sqrt{n_0}}G_{n0}(u+1-p) + o_p(1),
\end{aligned}$$

where

$$\begin{aligned}
G_{n1}(u) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n_1}\frac{\left[I\left\{F_1(Y_{1i}) \le u\right\} - u\right]}{f_1\left(F_1^{-1}(u)\right)}, \\
G_{n0}(u) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n_1}\frac{\left[I\left\{F_0(Y_{0i}) \le u\right\} - u\right]}{f_0\left(F_0^{-1}(1+u-p)\right)}.
\end{aligned}$$

Thus, we obtain

$$\begin{aligned}
& E\sup_{|u-u_{\sup,p}|<\eta}\left|\sqrt{n_1}\left(\Psi_n - \Psi\right)(u,p) - \sqrt{n_1}\left(\Psi_n - \Psi\right)(u_{\sup,p},p)\right| \\
& \quad \le E\sup_{|u-u_{\sup,p}|<\eta}\left|G_{n1}(u) - G_{n0}(u_{\sup,p})\right| \\
& \qquad + \sqrt{\lambda}E\sup_{|u-u_{\sup,p}|<\eta}\left|G_{n0}(u+1-p) - G_{n0}(u_{\sup,p}+1-p)\right|.
\end{aligned}$$

Assuming $\inf_u f_1\left(F_1^{-1}(u)\right) > 0$ and $\inf_u f_0\left(F_0^{-1}(u)\right) > 0$, we conclude that

$$E \sup_{|u-u_{\sup,p}|<\eta} \left|G_{n1}(u) - G_{n1}(u_{\sup,p})\right| \lesssim \eta^{1/2} \tag{IV.7}$$

and

$$E \sup_{|u-u_{\sup,p}|<\eta} \left|G_{n0}(1+u-p) - G_{n0}(1+u_{\sup,p}-p)\right| \lesssim \eta^{1/2}. \tag{IV.8}$$

Indeed, the envelope function of the class of functions

$$\{I\{(-\infty,u]\} - I\{(-\infty,u_{\sup,p}\} : u \in [u_{\sup,p}-\eta, u_{\sup,p}+\eta]\}$$

is bounded by $I\{(u_{\sup,p}-\eta, u_{\sup,p}+\eta)\}$ which has a squared L_2-norm bounded by 2η. Since the class of functions $I\{Y_{1i} \leq \cdot\}$ has a finite uniform entropy integral, Lemma 19.38 in van der Vaart (1998) implies the above results. Consequently,

$$E \sup_{|u-u_{\sup,p}|<\eta} |\sqrt{n_1}(\Psi_n - \Psi)(u,p) - \sqrt{n_1}(\Psi_n - \Psi)(u_{\sup,p},p)| \lesssim \eta^{1/2}$$

Hence the second condition of Theorem 3.2.5 in van der Vaart and Wellner (1996) is satisfied leading to the rate of $n_1^{-1/3}$.

Proof of 3. For a fixed p, we get

$$\begin{aligned}
&\sqrt{n_1}\left(Q_n^L(p) - Q^L(p)\right) \\
=\ & \sqrt{n_1}\left(F_{1n}^{-1}(\hat{u}_{\sup,p}) - F_{0n}^{-1}(1+\hat{u}_{\sup,p} - p)\right) \\
& -\sqrt{n_1}\left(F_1^{-1}(u_{\sup,p}) - F_0^{-1}(1+u_{\sup,p} - p)\right) \\
=\ & \sqrt{n_1}\left(\Psi_n - \Psi\right)(\hat{u}_{\sup,p}, p) + \sqrt{n_1}\Psi(\hat{u}_{\sup,p}, p) - \sqrt{n_1}\Psi(u_{\sup,p}, p) \\
=\ & \sqrt{n_1}\left(\Psi_n - \Psi\right)(u_{\sup,p}, p) \\
& +\sqrt{n_1}\left[\Psi(\hat{u}_{\sup,p}, p) - \Psi(u_{\sup,p}, p)\right] + o_p(1) \\
=\ & \sqrt{n_1}\left(\overline{Q}_n(p) - Q^L(p)\right) \\
& +\frac{1}{2}\sqrt{n_1}\left[\frac{\partial\Psi\left(u^*_{\sup,p}, p\right)}{\partial u}\right](\hat{u}_{\sup,p} - u_{\sup,p})^2 + o_p(1) \\
=\ & \sqrt{n_1}\left(\overline{Q}_n(p) - Q^L(p)\right) + o_p(1),
\end{aligned}$$

where $u^*_{\sup,p}$ lies between $\hat{u}_{\sup,p}$ and $u_{\sup,p}$ and we have used stochastic equicontinuity of the process:$\sqrt{n_1}(\Psi_n - \Psi)(\cdot, p)$ and the first order condition for $\sup \Psi(u, p)$.

<Proof for Lemma 7>

We prove (i) and (ii) is similar. Lemma 21.7 in van der Varrt (1998), along with

$f_1\left(F_1^{-1}\left(u_{\sup,p}\right)\right) = f_0\left(F_0^{-1}\left(1+u_{\sup,p}-p\right)\right)$, implies

$$
\begin{aligned}
& Y_{1\left(\left[n_1 u_{1nL}^A\right]\right)} - Y_{0\left(\left[n_0 u_{0nL}^A\right]\right)} \\
= \; & F_{1n}^{-1}\left(u_{\sup,p}\right) \\
& -z_{1-\alpha/2}\frac{u_{\sup,p}\left(1-u_{\sup,p}\right)}{\sqrt{n_1}\sqrt{u_{\sup,p}\left(1-u_{\sup,p}\right)+\frac{1}{\lambda}\left(1+u_{\sup,p}-p\right)\left(p-u_{\sup,p}\right)}f_1\left(F_1^{-1}\left(u_{\sup,p}\right)\right)} \\
& -F_{0n}^{-1}\left(1+u_{\sup,p}-p\right) \\
& -z_{1-\alpha/2}\frac{\left(1+u_{\sup,p}-p\right)\left(p-u_{\sup,p}\right)}{\sqrt{n_0}\sqrt{u_{\sup,p}\left(1-u_{\sup,p}\right)+\frac{1}{\lambda}\left(1+u_{\sup,p}-p\right)\left(p-u_{\sup,p}\right)}f_0\left(F_0^{-1}\left(1+u_{\sup,p}-p\right)\right)} \\
& +o_p\left(\frac{1}{\sqrt{n_1}}\right) \\
= \; & F_{1n}^{-1}\left(u_{\sup,p}\right) - F_{0n}^{-1}\left(1+u_{\sup,p}-p\right) \\
& -z_{1-\alpha/2}\frac{\sqrt{u_{\sup,p}\left(1-u_{\sup,p}\right)+\frac{1}{\lambda}\left(1+u_{\sup,p}-p\right)\left(p-u_{\sup,p}\right)}}{\sqrt{n_1}f_1\left(F_1^{-1}\left(u_{\sup,p}\right)\right)} + o_p\left(\frac{1}{\sqrt{n_1}}\right) \\
= \; & \overline{Q}_n(p) - z_{1-\alpha/2}\frac{\sigma_L}{\sqrt{n_1}} + o_p\left(\frac{1}{\sqrt{n_1}}\right),
\end{aligned}
$$

where $\overline{Q}_n(p) = F_{1n}^{-1}\left(u_{\sup,p}\right) - F_{0n}^{-1}\left(1+u_{\sup,p}-p\right)$.

Similarly, we obtain

$$
Y_{1\left(\left[n_1 u_{1nL}^B\right]\right)} - Y_{0\left(\left[n_0 u_{0nL}^B\right]\right)} = \overline{Q}_n(p) + z_{1-\alpha/2}\frac{\sigma_L}{\sqrt{n_1}} + o_p\left(\frac{1}{\sqrt{n_1}}\right).
$$

The conclusion in (i) follows from: $\sqrt{n_1}\left(\overline{Q}_n\left(p\right) - Q^L\left(p\right)\right) \Longrightarrow N\left(0,\sigma_L^2\right)$.

<**Proof for Theorem 6**> By the stochastic equicontinuity of the quantile

processes and the proof of Lemma 7, we obtain:

$$
\begin{aligned}
& Y_{1\left(\left[n_1 \hat{u}_{1nL}^A\right]\right)} - Y_{0\left(\left[n_0 \hat{u}_{0nL}^A\right]\right)} \\
= \;& F_{1n}^{-1}\left(\hat{u}_{1nL}^A\right) - F_{0n}^{-1}\left(\hat{u}_{0nL}^A\right) \\
= \;& \left\{\left[F_{1n}^{-1}\left(\hat{u}_{1nL}^A\right) - F_{0n}^{-1}\left(\hat{u}_{0nL}^A\right)\right] - \left[F_1^{-1}\left(\hat{u}_{1nL}^A\right) - F_0^{-1}\left(\hat{u}_{0nL}^A\right)\right]\right\} \\
& + \left[F_1^{-1}\left(\hat{u}_{1nL}^A\right) - F_0^{-1}\left(\hat{u}_{0nL}^A\right)\right] \\
= \;& \left[F_{1n}^{-1}\left(u_{1nL}^A\right) - F_{0n}^{-1}\left(u_{0nL}^A\right)\right] - \left[F_1^{-1}\left(u_{1nL}^A\right) - F_0^{-1}\left(u_{0nL}^A\right)\right] \\
& + \left[F_1^{-1}\left(\hat{u}_{1nL}^A\right) - F_0^{-1}\left(\hat{u}_{0nL}^A\right)\right] + o_p\left(\frac{1}{\sqrt{n_1}}\right) \\
= \;& \left[Y_{1\left(\left[n_1 u_{1nL}^A\right]\right)} - Y_{0\left(\left[n_0 u_{0nL}^A\right]\right)}\right] \\
& + \left\{\left[F_1^{-1}\left(\hat{u}_{1nL}^A\right) - F_0^{-1}\left(\hat{u}_{0nL}^A\right)\right] - \left[F_1^{-1}\left(u_{1nL}^A\right) - F_0^{-1}\left(u_{0nL}^A\right)\right]\right\} + o_p\left(\frac{1}{\sqrt{n_1}}\right) \\
= \;& \overline{Q}_n(p) - z_{1-\alpha/2}\frac{\sigma_L}{\sqrt{n_1}} \\
& + \left\{\left[F_1^{-1}\left(\hat{u}_{1nL}^A\right) - F_0^{-1}\left(\hat{u}_{0nL}^A\right)\right] - \left[F_1^{-1}\left(u_{1nL}^A\right) - F_0^{-1}\left(u_{0nL}^A\right)\right]\right\} + o_p\left(\frac{1}{\sqrt{n_1}}\right) \\
= \;& \overline{Q}_n(p) - z_{1-\alpha/2}\frac{\sigma_L}{\sqrt{n_1}} + o_p\left(\frac{1}{\sqrt{n_1}}\right),
\end{aligned}
$$

where $\left[F_1^{-1}\left(\hat{u}_{1nL}^A\right) - F_0^{-1}\left(\hat{u}_{0nL}^A\right)\right] - \left[F_1^{-1}\left(u_{1nL}^A\right) - F_0^{-1}\left(u_{0nL}^A\right)\right] = o_p\left(n_1^{-1/2}\right)$, by Taylor series expansion, $f_1\left(F_1^{-1}\left(u_{\sup,p}\right)\right) = f_0\left(F_0^{-1}\left(1 + u_{\sup,p} - p\right)\right)$, and $\hat{u}_{\sup,p} - u_{\sup,p} = O_p\left(n_1^{-1/3}\right)$. Similarly, we obtain

$$
Y_{1\left(\left[n_1 \hat{u}_{1nL}^B\right]\right)} - Y_{0\left(\left[n_0 \hat{u}_{0nL}^B\right]\right)} = \overline{Q}_n(p) + z_{1-\alpha/2}\frac{\sigma_L}{\sqrt{n_1}} + o_p\left(\frac{1}{\sqrt{n_1}}\right).
$$

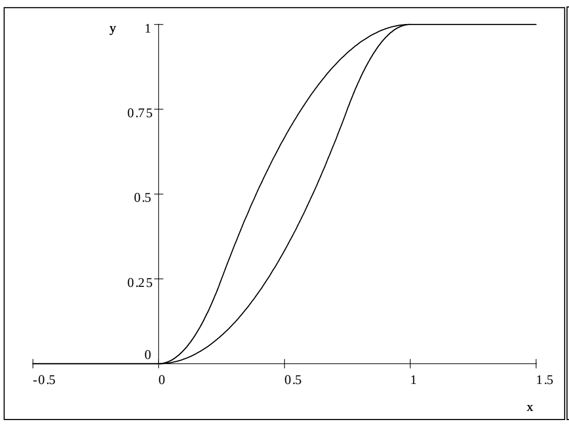
The conclusion in (i) follows from: $\sqrt{n_1}\left(\overline{Q}_n(p) - Q^L(p)\right) \Longrightarrow N\left(0, \sigma_L^2\right)$.

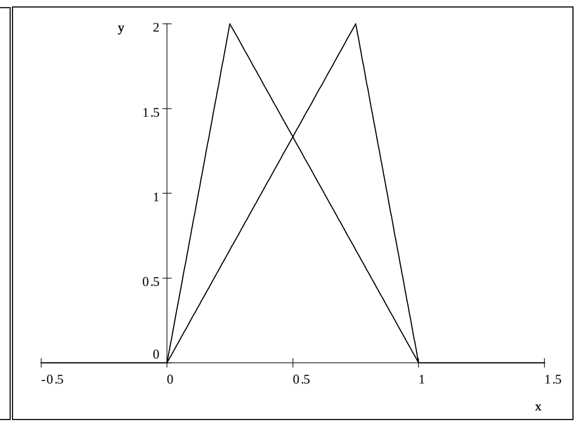
Appendix B. Graphs of Data Generating Processes

<Model 1> $Y_1 \sim N(2,2)$ and $Y_0 \sim N(1,1)$

<Model 2> $Y_1 \sim C\left(\frac{1}{4}\right)$ and $Y_0 \sim C\left(\frac{3}{4}\right)$ where

$$X \sim C(a) \implies P(X \le x) = \begin{cases} \frac{1}{a}x^2 \text{ if } x \in [0,a] \\ 1 - \frac{(x-1)^2}{(1-a)} \text{ if } x \in [a,1] \end{cases}$$





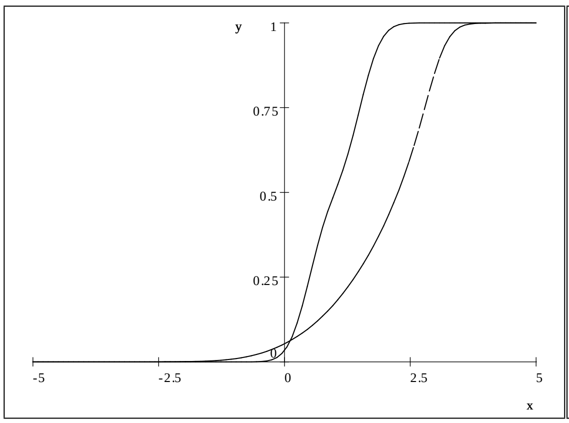
Solid line: $F_0(x)$, dashed line: $F_1(x)$

Associated $f_0(x)$ and $f_1(x)$

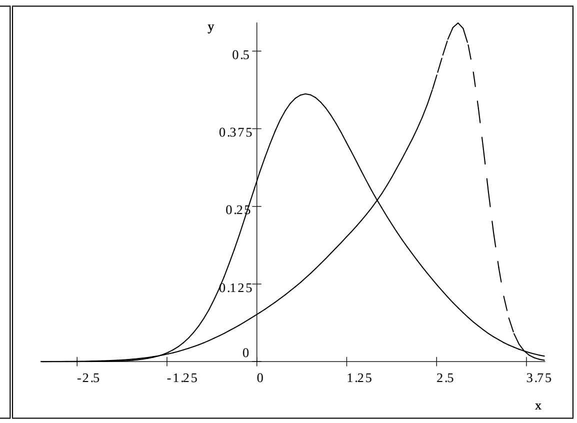
<Model 3> Two skewed unimodal distributions

$$F_1(x) = \frac{1}{4}\sum_{i=0}^{3} N\left(3\left(1-\left(\frac{2}{3}\right)^i\right)+\frac{29}{36}, \left(\frac{2}{3}\right)^{2i}\right),$$

$$F_0(x) = \frac{1}{2}\sum_{i=0}^{1} N\left(3\left(\left(\frac{2}{3}\right)^i-1\right)+\frac{3}{2}, \left(\frac{2}{3}\right)^{2i}\right).$$



Solid line: $F_0(x)$, dashed line: $F_1(x)$

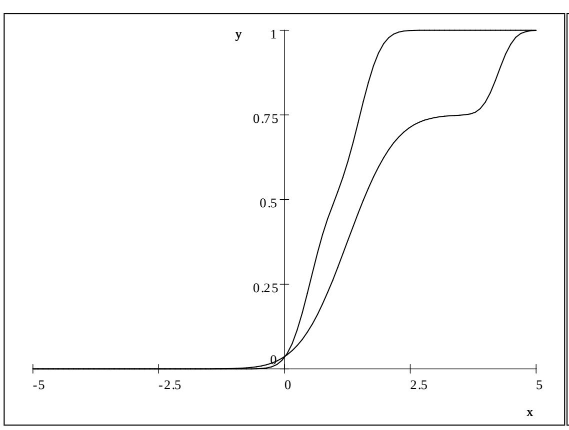


Associated $f_0(x)$ and $f_1(x)$

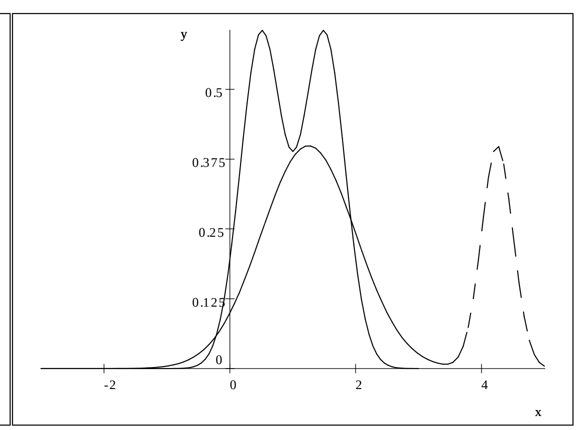
<Model 4> Two bimodal distributions

$$F_1(x) = \frac{3}{4}N\left(\frac{5}{4}, \left(\frac{3}{4}\right)^2\right)+\frac{1}{4}N\left(\frac{17}{4}, \left(\frac{1}{4}\right)^2\right),$$

$$F_0(x) = \frac{1}{2}N\left(\frac{1}{2}, \left(\frac{1}{3}\right)^2\right)+\frac{1}{2}N\left(\frac{3}{2}, \left(\frac{1}{3}\right)^2\right).$$



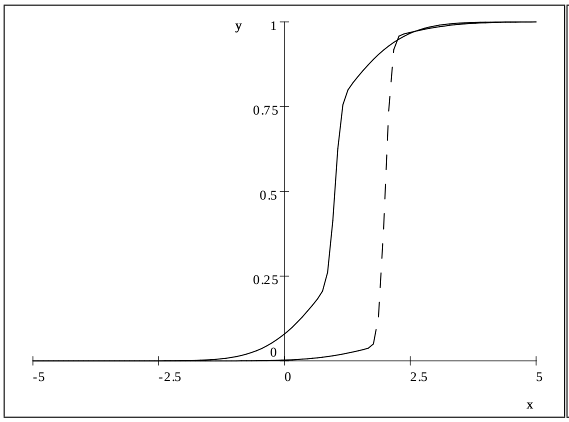
Solid line: $F_0(x)$, dashed line: $F_1(x)$

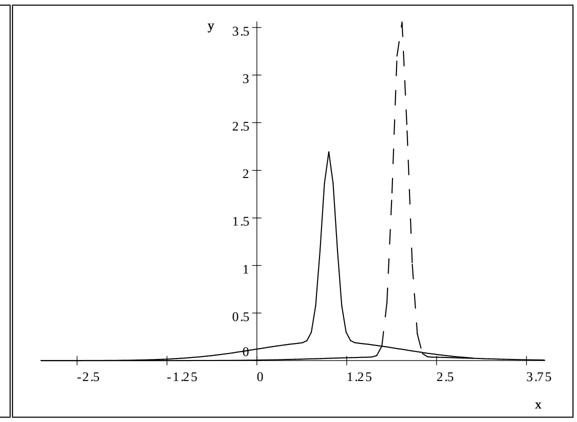


Associated $f_0(x)$ and $f_1(x)$

<Model 5> Two kurtotic distributions

$$F_1(x) = \frac{1}{10}N(2,1) + \frac{9}{10}N\left(2, \left(\frac{1}{10}\right)^2\right),$$
$$F_0(x) = \frac{1}{2}N(1,1) + \frac{1}{2}N\left(1, \left(\frac{1}{10}\right)^2\right).$$



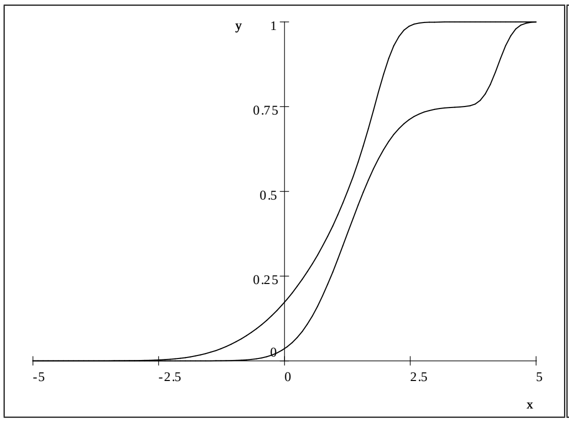


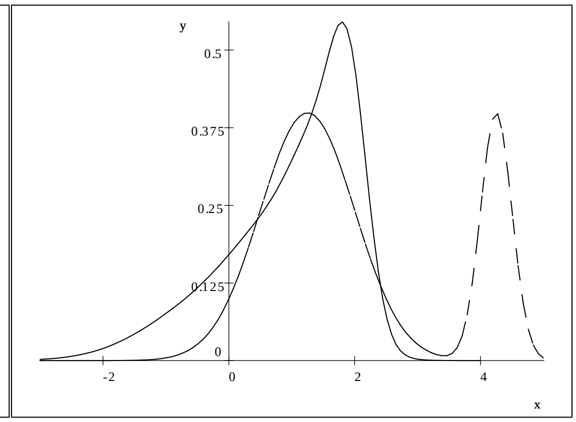
Solid line: $F_0(x)$, dashed line: $F_1(x)$ Associated $f_0(x)$ and $f_1(x)$

<Model 6> Y_0 follows a skewed unimodal distribution and y_1 a bimodal distribu-

tion

$$F_1(x) = \frac{3}{4} N\left(\frac{5}{4}, \left(\frac{3}{4}\right)^2\right) + \frac{1}{4} N\left(\frac{17}{4}, \left(\frac{1}{4}\right)^2\right),$$
$$F_0(x) = \frac{1}{4}\sum_{i=0}^{3} N\left(3\left(1-\left(\frac{2}{3}\right)^i\right) - \frac{7}{36}, \left(\frac{2}{3}\right)^{2i}\right).$$





Solid line: $F_0(x)$, dashed line: $F_1(x)$ Associated $f_0(x)$ and $f_1(x)$

Appendix C. Tables of simulation results for Each Bound

<Model 1>

		p=0.1 C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)	p=0.3 C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)
n=1000	New	0.943	0.965	0.4586	0.3211	0.951	0.958	0.3339	0.2880
	$A(h_S)$	0.942	0.942	0.4669	0.3248	0.957	0.938	0.3396	0.2922
	$A(h_R)$	0.942	0.942	0.4671	0.3251	0.957	0.938	0.3399	0.2924
	$A(h_M)$	0.937	0.939	0.4620	0.3219	0.954	0.936	0.3364	0.2895
n=2000	New	0.954	0.960	0.3241	0.2245	0.952	0.958	0.2346	0.2011
	$A(h_S)$	0.945	0.926	0.3266	0.2272	0.948	0.943	0.2377	0.2050
	$A(h_R)$	0.944	0.927	0.3267	0.2273	0.948	0.943	0.2379	0.2051
	$A(h_M)$	0.943	0.925	0.3242	0.2256	0.948	0.937	0.2360	0.2035
n=4000	New	0.958	0.956	0.2266	0.1578	0.952	0.954	0.1654	0.1426
	$A(h_S)$	0.951	0.943	0.2295	0.1593	0.940	0.945	0.1668	0.1439
	$A(h_R)$	0.951	0.943	0.2296	0.1594	0.940	0.945	0.1669	0.1440
	$A(h_M)$	0.949	0.942	0.2282	0.1585	0.940	0.942	0.1659	0.1431
n=6000	New	0.952	0.957	0.1849	0.1284	0.943	0.959	0.1348	0.1161
	$A(h_S)$	0.950	0.941	0.1869	0.1297	0.932	0.950	0.1359	0.1173
	$A(h_R)$	0.950	0.941	0.1869	0.1297	0.932	0.950	0.1359	0.1173
	$A(h_M)$	0.949	0.941	0.1860	0.1291	0.931	0.948	0.1352	0.1167

		p=0.5 C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)	p=0.7 C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)
n=1000	New	0.956	0.958	0.2973	0.2970	0.949	0.958	0.2857	0.3318
	$A(h_S)$	0.951	0.944	0.3039	0.3031	0.945	0.941	0.2923	0.3392
	$A(h_R)$	0.951	0.944	0.3040	0.3033	0.944	0.941	0.2925	0.3395
	$A(h_M)$	0.948	0.943	0.3010	0.3002	0.942	0.939	0.2896	0.3359
n=2000	New	0.951	0.951	0.2088	0.2087	0.951	0.944	0.2015	0.2344
	$A(h_S)$	0.942	0.944	0.2128	0.2124	0.946	0.928	0.2051	0.2375
	$A(h_R)$	0.942	0.942	0.2128	0.2125	0.946	0.928	0.2052	0.2376
	$A(h_M)$	0.941	0.939	0.2112	0.2108	0.945	0.924	0.2037	0.2358
n=4000	New	0.961	0.947	0.1474	0.1470	0.967	0.960	0.1421	0.1650
	$A(h_S)$	0.948	0.954	0.1495	0.1493	0.942	0.955	0.1439	0.1670
	$A(h_R)$	0.948	0.954	0.1495	0.1493	0.942	0.955	0.1439	0.1670
	$A(h_M)$	0.946	0.954	0.1487	0.1484	0.941	0.954	0.1431	0.1661
n=6000	New	0.954	0.953	0.1201	0.1207	0.939	0.958	0.1159	0.1346
	$A(h_S)$	0.942	0.950	0.1217	0.1215	0.935	0.943	0.1172	0.1359
	$A(h_R)$	0.941	0.950	0.1217	0.1216	0.935	0.943	0.1172	0.1359
	$A(h_M)$	0.940	0.950	0.1211	0.1210	0.934	0.942	0.1166	0.1353

		p=0.9			
		C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)
n=1000	New	0.969	0.959	0.3205	0.4678
	$A\,(h_S)$	0.943	0.944	0.3243	0.4689
	$A\,(h_R)$	0.943	0.943	0.3245	0.4691
	$A\,(h_M)$	0.941	0.943	0.3214	0.4644
n=2000	New	0.945	0.964	0.2232	0.3256
	$A\,(h_S)$	0.939	0.950	0.2270	0.3276
	$A\,(h_R)$	0.939	0.950	0.2271	0.3277
	$A\,(h_M)$	0.936	0.949	0.2254	0.3252
n=4000	New	0.935	0.961	0.1572	0.2279
	$A\,(h_S)$	0.945	0.949	0.1592	0.2294
	$A\,(h_R)$	0.945	0.949	0.1593	0.2295
	$A\,(h_M)$	0.941	0.947	0.1584	0.2281
n=6000	New	0.939	0.959	0.1285	0.1850
	$A\,(h_S)$	0.930	0.949	0.1295	0.1863
	$A\,(h_R)$	0.930	0.949	0.1296	0.1864
	$A\,(h_M)$	0.926	0.946	0.1289	0.1854

<Model 2>

		p=0.1 C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)	p=0.3 C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)
n=1000	New	0.943	0.965	0.4586	0.3211	0.951	0.958	0.3339	0.2880
	A(h_S)	0.942	0.942	0.4669	0.3248	0.957	0.938	0.3396	0.2922
	A(h_R)	0.942	0.942	0.4671	0.3251	0.957	0.938	0.3399	0.2924
	A(h_M)	0.937	0.939	0.4620	0.3219	0.954	0.936	0.3364	0.2895
n=2000	New	0.954	0.960	0.3241	0.2245	0.952	0.958	0.2346	0.2011
	A(h_S)	0.945	0.926	0.3266	0.2272	0.948	0.943	0.2377	0.2050
	A(h_R)	0.944	0.927	0.3267	0.2273	0.948	0.943	0.2379	0.2051
	A(h_M)	0.943	0.925	0.3242	0.2256	0.948	0.937	0.2360	0.2035
n=4000	New	0.958	0.956	0.2266	0.1578	0.952	0.954	0.1654	0.1426
	A(h_S)	0.951	0.943	0.2295	0.1593	0.940	0.945	0.1668	0.1439
	A(h_R)	0.951	0.943	0.2296	0.1594	0.940	0.945	0.1669	0.1440
	A(h_M)	0.949	0.942	0.2282	0.1585	0.940	0.942	0.1659	0.1431
n=6000	New	0.952	0.957	0.1849	0.1284	0.943	0.959	0.1348	0.1161
	A(h_S)	0.950	0.941	0.1869	0.1297	0.932	0.950	0.1359	0.1173
	A(h_R)	0.950	0.941	0.1869	0.1297	0.932	0.950	0.1359	0.1173
	A(h_M)	0.949	0.941	0.1860	0.1291	0.931	0.948	0.1352	0.1167

		p=0.5 C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)	p=0.7 C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)
n=1000	New	0.956	0.958	0.2973	0.2970	0.949	0.958	0.2857	0.3318
	A(h_S)	0.951	0.944	0.3039	0.3031	0.945	0.941	0.2923	0.3392
	A(h_R)	0.951	0.944	0.3040	0.3033	0.944	0.941	0.2925	0.3395
	A(h_M)	0.948	0.943	0.3010	0.3002	0.942	0.939	0.2896	0.3359
n=2000	New	0.951	0.951	0.2088	0.2087	0.951	0.944	0.2015	0.2344
	A(h_S)	0.942	0.944	0.2128	0.2124	0.946	0.928	0.2051	0.2375
	A(h_R)	0.942	0.942	0.2128	0.2125	0.946	0.928	0.2052	0.2376
	A(h_M)	0.941	0.939	0.2112	0.2108	0.945	0.924	0.2037	0.2358
n=4000	New	0.961	0.947	0.1474	0.1470	0.967	0.960	0.1421	0.1650
	A(h_S)	0.948	0.954	0.1495	0.1493	0.942	0.955	0.1439	0.1670
	A(h_R)	0.948	0.954	0.1495	0.1493	0.942	0.955	0.1439	0.1670
	A(h_M)	0.946	0.954	0.1487	0.1484	0.941	0.954	0.1431	0.1661
n=6000	New	0.954	0.953	0.1201	0.1207	0.939	0.958	0.1159	0.1346
	A(h_S)	0.942	0.950	0.1217	0.1215	0.935	0.943	0.1172	0.1359
	A(h_R)	0.941	0.950	0.1217	0.1216	0.935	0.943	0.1172	0.1359
	A(h_M)	0.940	0.950	0.1211	0.1210	0.934	0.942	0.1166	0.1353

		p=0.9			
		C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)
n=1000	New	0.969	0.959	0.3205	0.4678
	A(h_S)	0.943	0.944	0.3243	0.4689
	A(h_R)	0.943	0.943	0.3245	0.4691
	A(h_M)	0.941	0.943	0.3214	0.4644
n=2000	New	0.945	0.964	0.2232	0.3256
	A(h_S)	0.939	0.950	0.2270	0.3276
	A(h_R)	0.939	0.950	0.2271	0.3277
	A(h_M)	0.936	0.949	0.2254	0.3252
n=4000	New	0.935	0.961	0.1572	0.2279
	A(h_S)	0.945	0.949	0.1592	0.2294
	A(h_R)	0.945	0.949	0.1593	0.2295
	A(h_M)	0.941	0.947	0.1584	0.2281
n=6000	New	0.939	0.959	0.1285	0.1850
	A(h_S)	0.930	0.949	0.1295	0.1863
	A(h_R)	0.930	0.949	0.1296	0.1864
	A(h_M)	0.926	0.946	0.1289	0.1854

<Model 3>

		p=0.1 C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)	p=0.3 C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)
n=1000	New	0.953	0.959	0.4923	0.2857	0.955	0.956	0.3736	0.1927
	A(h_S)	0.939	0.961	0.4955	0.3002	0.945	0.947	0.3707	0.2092
	A(h_R)	0.938	0.962	0.4956	0.3005	0.945	0.946	0.3710	0.2093
	A(h_M)	0.934	0.957	0.4922	0.2947	0.945	0.941	0.3697	0.2041
n=2000	New	0.950	0.952	0.3468	0.1938	0.945	0.956	0.2611	0.1367
	A(h_S)	0.941	0.955	0.3473	0.2048	0.948	0.947	0.2613	0.1452
	A(h_R)	0.941	0.954	0.3473	0.2050	0.948	0.948	0.2614	0.1453
	A(h_M)	0.939	0.950	0.3455	0.2017	0.948	0.940	0.2607	0.1425
n=4000	New	0.959	0.975	0.2442	0.1370	0.949	0.961	0.1843	0.0962
	A(h_S)	0.946	0.971	0.2447	0.1417	0.943	0.961	0.1841	0.1009
	A(h_R)	0.946	0.972	0.2448	0.1418	0.943	0.961	0.1842	0.1010
	A(h_M)	0.944	0.966	0.2438	0.1400	0.943	0.959	0.1838	0.0995
n=6000	New	0.960	0.948	0.1990	0.1112	0.944	0.946	0.1508	0.0784
	A(h_S)	0.946	0.947	0.1995	0.1148	0.925	0.950	0.1502	0.0818
	A(h_R)	0.946	0.948	0.1995	0.1148	0.925	0.950	0.1502	0.0818
	A(h_M)	0.946	0.945	0.1988	0.1136	0.925	0.944	0.1500	0.0807

		p=0.5 C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)	p=0.7 C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)
n=1000	New	0.955	0.950	0.3168	0.1704	0.948	0.954	0.2705	0.1764
	A(h_S)	0.939	0.966	0.3132	0.1891	0.925	0.962	0.2686	0.1951
	A(h_R)	0.939	0.965	0.3132	0.1892	0.925	0.961	0.2685	0.1953
	A(h_M)	0.939	0.963	0.3129	0.1835	0.926	0.947	0.2680	0.1888
n=2000	New	0.954	0.957	0.2228	0.1207	0.966	0.962	0.1903	0.1226
	A(h_S)	0.947	0.940	0.2213	0.1304	0.949	0.963	0.1898	0.1342
	A(h_R)	0.947	0.940	0.2213	0.1304	0.949	0.963	0.1897	0.1343
	A(h_M)	0.945	0.933	0.2211	0.1273	0.950	0.956	0.1895	0.1308
n=4000	New	0.959	0.962	0.1573	0.0847	0.959	0.954	0.1338	0.0865
	A(h_S)	0.945	0.961	0.1563	0.0901	0.946	0.955	0.1339	0.0926
	A(h_R)	0.945	0.961	0.1563	0.0901	0.946	0.956	0.1338	0.0927
	A(h_M)	0.944	0.957	0.1563	0.0884	0.945	0.954	0.1337	0.0907
n=6000	New	0.944	0.950	0.1283	0.0688	0.945	0.953	0.1094	0.0703
	A(h_S)	0.926	0.949	0.1276	0.0728	0.934	0.958	0.1093	0.0748
	A(h_R)	0.926	0.949	0.1276	0.0728	0.934	0.957	0.1092	0.0748
	A(h_M)	0.924	0.943	0.1276	0.0716	0.933	0.954	0.1092	0.0735

		p=0.9			
		C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)
n=1000	New	0.956	0.959	0.2344	0.2257
	A(h_S)	0.946	0.963	0.2343	0.2521
	A(h_R)	0.946	0.965	0.2342	0.2524
	A(h_M)	0.945	0.959	0.2331	0.2433
n=2000	New	0.951	0.957	0.1640	0.1574
	A(h_S)	0.944	0.960	0.1648	0.1728
	A(h_R)	0.944	0.960	0.1648	0.1729
	A(h_M)	0.943	0.949	0.1642	0.1680
n=4000	New	0.959	0.945	0.1162	0.1109
	A(h_S)	0.947	0.953	0.1161	0.1189
	A(h_R)	0.947	0.955	0.1161	0.1190
	A(h_M)	0.947	0.950	0.1159	0.1164
n=6000	New	0.951	0.958	0.0948	0.0900
	A(h_S)	0.940	0.963	0.0946	0.0958
	A(h_R)	0.940	0.963	0.0946	0.0959
	A(h_M)	0.939	0.960	0.0945	0.0940

<Model 4>

		p=0.1				p=0.3			
		$C(Q^L)$	$C(Q^U)$	$W(Q^L)$	$W(Q^U)$	$C(Q^L)$	$C(Q^U)$	$W(Q^L)$	$W(Q^U)$
n=1000	New	0.954	0.972	0.2267	0.1911	0.957	0.950	0.1788	0.1799
	$A(h_S)$	0.979	0.976	0.2712	0.2246	0.973	0.979	0.2128	0.2144
	$A(h_R)$	0.976	0.973	0.2666	0.2207	0.972	0.971	0.2089	0.2100
	$A(h_M)$	0.967	0.959	0.2522	0.2089	0.963	0.961	0.1979	0.1992
n=2000	New	0.941	0.950	0.1609	0.1340	0.950	0.947	0.1261	0.1264
	$A(h_S)$	0.969	0.965	0.1839	0.1529	0.976	0.967	0.1445	0.1454
	$A(h_R)$	0.969	0.961	0.1817	0.1510	0.973	0.962	0.1426	0.1434
	$A(h_M)$	0.963	0.952	0.1738	0.1446	0.965	0.952	0.1366	0.1374
n=4000	New	0.949	0.957	0.1133	0.0944	0.962	0.948	0.0891	0.0894
	$A(h_S)$	0.974	0.966	0.1258	0.1047	0.961	0.964	0.0989	0.0996
	$A(h_R)$	0.972	0.962	0.1248	0.1039	0.959	0.963	0.0982	0.0987
	$A(h_M)$	0.965	0.957	0.1205	0.1003	0.957	0.957	0.0948	0.0953
n=6000	New	0.952	0.963	0.0922	0.0766	0.943	0.960	0.0724	0.0729
	$A(h_S)$	0.962	0.969	0.1011	0.0840	0.958	0.965	0.0795	0.0799
	$A(h_R)$	0.963	0.967	0.1005	0.0835	0.956	0.962	0.0790	0.0794
	$A(h_M)$	0.952	0.959	0.0975	0.0810	0.948	0.957	0.0767	0.0770

		p=0.5				p=0.7			
		$C(Q^L)$	$C(Q^U)$	$W(Q^L)$	$W(Q^U)$	$C(Q^L)$	$C(Q^U)$	$W(Q^L)$	$W(Q^U)$
n=1000	New	0.954	0.947	0.1811	0.2214	0.954	0.947	0.3659	0.8295
	$A(h_S)$	0.972	0.971	0.2161	0.2652	0.986	0.943	0.4134	0.6311
	$A(h_R)$	0.970	0.965	0.2118	0.2589	0.982	0.928	0.4028	0.6268
	$A(h_M)$	0.965	0.961	0.2009	0.2458	0.983	0.929	0.3927	0.6572
n=2000	New	0.952	0.945	0.1279	0.1568	0.940	0.955	0.2459	0.4899
	$A(h_S)$	0.967	0.969	0.1468	0.1803	0.984	0.953	0.2829	0.4685
	$A(h_R)$	0.967	0.962	0.1447	0.1770	0.980	0.940	0.2775	0.4624
	$A(h_M)$	0.959	0.955	0.1387	0.1698	0.977	0.934	0.2679	0.4670
n=4000	New	0.951	0.945	0.0900	0.1097	0.944	0.951	0.1704	0.3121
	$A(h_S)$	0.960	0.972	0.1005	0.1231	0.969	0.969	0.1917	0.3328
	$A(h_R)$	0.958	0.966	0.0996	0.1217	0.967	0.958	0.1891	0.3286
	$A(h_M)$	0.953	0.959	0.0962	0.1177	0.963	0.959	0.1832	0.3229
n=6000	New	0.928	0.952	0.0735	0.0893	0.936	0.951	0.1393	0.2515
	$A(h_S)$	0.948	0.968	0.0807	0.0988	0.966	0.966	0.1542	0.2707
	$A(h_R)$	0.944	0.962	0.0802	0.0979	0.962	0.960	0.1524	0.2675
	$A(h_M)$	0.943	0.960	0.0778	0.0950	0.956	0.958	0.1482	0.2618

		p=0.9 C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)
n=1000	New	0.951	0.884	0.1702	2.4675
	A(h_S)	1.000	1.000	0.3972	3.3368
	A(h_R)	1.000	1.000	0.3573	7.9198
	A(h_M)	1.000	1.000	0.3142	26.2168
n=2000	New	0.953	0.886	0.1191	2.2881
	A(h_S)	1.000	1.000	0.2526	3.6706
	A(h_R)	1.000	1.000	0.2306	8.4325
	A(h_M)	0.999	1.000	0.2038	32.1941
n=4000	New	0.949	0.887	0.0830	2.1670
	A(h_S)	1.000	1.000	0.1594	4.8924
	A(h_R)	0.999	1.000	0.1484	11.4761
	A(h_M)	0.999	1.000	0.1321	53.1956
n=6000	New	0.942	0.897	0.0677	2.1125
	A(h_S)	1.000	1.000	0.1230	5.8978
	A(h_R)	1.000	1.000	0.1155	16.1382
	A(h_M)	0.998	1.000	0.1034	96.5903

<Model 5>

		p=0.1 C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)	p=0.3 C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)
n=1000	New	0.940	0.946	0.4506	0.3348	0.937	0.961	0.1141	0.0368
	A(h_S)	0.929	0.926	0.3658	0.2786	0.968	0.999	0.1183	0.0629
	A(h_R)	0.921	0.940	0.3845	0.3214	0.946	0.970	0.1072	0.0406
	A(h_M)	0.920	0.942	0.3838	0.3261	0.943	0.962	0.1071	0.0395
n=2000	New	0.926	0.963	0.3031	0.2243	0.942	0.950	0.0763	0.0257
	A(h_S)	0.923	0.930	0.2591	0.2042	0.971	0.998	0.0815	0.0424
	A(h_R)	0.923	0.942	0.2727	0.2180	0.958	0.942	0.0746	0.0279
	A(h_M)	0.922	0.944	0.2731	0.2186	0.959	0.938	0.0744	0.0273
n=4000	New	0.947	0.965	0.2081	0.1532	0.960	0.973	0.0523	0.0182
	A(h_S)	0.932	0.948	0.1836	0.1448	0.967	0.995	0.0552	0.0268
	A(h_R)	0.932	0.956	0.1938	0.1507	0.949	0.972	0.0516	0.0193
	A(h_M)	0.930	0.953	0.1942	0.1508	0.948	0.968	0.0516	0.0189
n=6000	New	0.952	0.930	0.1673	0.1239	0.938	0.950	0.0424	0.0148
	A(h_S)	0.934	0.928	0.1515	0.1205	0.954	0.991	0.0453	0.0219
	A(h_R)	0.945	0.935	0.1588	0.1227	0.938	0.950	0.0420	0.0156
	A(h_M)	0.947	0.934	0.1592	0.1228	0.938	0.947	0.0419	0.0153

		p=0.5 C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)	p=0.7 C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)
n=1000	New	0.952	0.958	0.0431	0.0433	0.967	0.958	0.0369	0.1147
	A(h_S)	0.998	0.999	0.0708	0.0708	0.997	0.973	0.0629	0.1182
	A(h_R)	0.962	0.971	0.0474	0.0474	0.963	0.940	0.0406	0.1070
	A(h_M)	0.959	0.966	0.0461	0.0462	0.956	0.938	0.0395	0.1069
n=2000	New	0.952	0.954	0.0304	0.0303	0.950	0.961	0.0257	0.0760
	A(h_S)	0.998	0.997	0.0473	0.0479	0.999	0.975	0.0418	0.0823
	A(h_R)	0.964	0.958	0.0326	0.0326	0.963	0.949	0.0279	0.0738
	A(h_M)	0.961	0.947	0.0320	0.0319	0.959	0.949	0.0272	0.0737
n=4000	New	0.955	0.962	0.0213	0.0214	0.963	0.955	0.0181	0.0520
	A(h_S)	0.994	0.997	0.0306	0.0306	0.998	0.969	0.0268	0.0552
	A(h_R)	0.961	0.962	0.0226	0.0226	0.961	0.953	0.0193	0.0515
	A(h_M)	0.957	0.956	0.0222	0.0222	0.957	0.952	0.0190	0.0514
n=6000	New	0.945	0.958	0.0173	0.0173	0.957	0.945	0.0148	0.0421
	A(h_S)	0.991	0.993	0.0250	0.0251	0.992	0.964	0.0219	0.0455
	A(h_R)	0.947	0.948	0.0182	0.0183	0.959	0.950	0.0156	0.0419
	A(h_M)	0.945	0.945	0.0180	0.0180	0.957	0.950	0.0153	0.0418

		p=0.9 $C(Q^L)$	$C(Q^U)$	$W(Q^L)$	$W(Q^U)$
n=1000	New	0.953	0.938	0.3372	0.4457
	$A(h_S)$	0.923	0.935	0.2697	0.3793
	$A(h_R)$	0.940	0.941	0.3014	0.5521
	$A(h_M)$	0.940	0.937	0.3013	0.9053
n=2000	New	0.935	0.951	0.2223	0.3047
	$A(h_S)$	0.909	0.946	0.1972	0.2674
	$A(h_R)$	0.925	0.954	0.2113	0.2934
	$A(h_M)$	0.925	0.955	0.2113	0.2984
n=4000	New	0.954	0.934	0.1535	0.2077
	$A(h_S)$	0.941	0.920	0.1441	0.1856
	$A(h_R)$	0.948	0.925	0.1492	0.1994
	$A(h_M)$	0.948	0.925	0.1492	0.2007
n=6000	New	0.941	0.959	0.1239	0.1662
	$A(h_S)$	0.933	0.937	0.1198	0.1516
	$A(h_R)$	0.936	0.941	0.1223	0.1603
	$A(h_M)$	0.938	0.944	0.1223	0.1610

<Model 6>

		p=0.1 C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)	p=0.3 C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)
n=1000	New	0.956	0.954	0.2301	0.2610	0.959	0.962	0.1812	0.3280
	A(h_S)	0.976	0.957	0.2688	0.2752	0.970	0.961	0.2099	0.3462
	A(h_R)	0.982	0.957	0.2794	0.2845	0.977	0.966	0.2181	0.3582
	A(h_M)	0.968	0.952	0.2575	0.2703	0.965	0.959	0.2013	0.3396
n=2000	New	0.943	0.951	0.1627	0.1839	0.947	0.943	0.1274	0.2300
	A(h_S)	0.964	0.946	0.1831	0.1913	0.973	0.951	0.1433	0.2400
	A(h_R)	0.973	0.948	0.1893	0.1967	0.981	0.955	0.1481	0.2470
	A(h_M)	0.963	0.942	0.1770	0.1888	0.969	0.947	0.1387	0.2367
n=4000	New	0.950	0.963	0.1147	0.1294	0.958	0.948	0.0898	0.1616
	A(h_S)	0.973	0.956	0.1256	0.1338	0.959	0.958	0.0985	0.1671
	A(h_R)	0.977	0.960	0.1293	0.1370	0.964	0.960	0.1014	0.1713
	A(h_M)	0.965	0.952	0.1223	0.1325	0.957	0.955	0.0960	0.1653
n=6000	New	0.956	0.947	0.0934	0.1057	0.942	0.948	0.0731	0.1317
	A(h_S)	0.962	0.950	0.1011	0.1086	0.954	0.956	0.0793	0.1354
	A(h_R)	0.965	0.952	0.1038	0.1110	0.960	0.957	0.0814	0.1384
	A(h_M)	0.956	0.950	0.0988	0.1077	0.950	0.956	0.0775	0.1342

		p=0.5 C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)	p=0.7 C(Q^L)	C(Q^U)	W(Q^L)	W(Q^U)
n=1000	New	0.951	0.953	0.1827	0.4510	0.949	0.967	0.3574	0.9214
	A(h_S)	0.969	0.954	0.2103	0.4766	0.982	0.932	0.3921	0.7380
	A(h_R)	0.973	0.957	0.2189	0.4889	0.980	0.931	0.4015	0.7035
	A(h_M)	0.961	0.950	0.2020	0.4662	0.979	0.925	0.3790	0.7568
n=2000	New	0.948	0.949	0.1286	0.3111	0.941	0.969	0.2416	0.6469
	A(h_S)	0.964	0.953	0.1436	0.3270	0.974	0.940	0.2651	0.5224
	A(h_R)	0.969	0.955	0.1487	0.3371	0.980	0.942	0.2759	0.5087
	A(h_M)	0.959	0.949	0.1392	0.3215	0.967	0.930	0.2578	0.5204
n=4000	New	0.946	0.951	0.0907	0.2192	0.941	0.973	0.1678	0.4099
	A(h_S)	0.952	0.960	0.0989	0.2282	0.964	0.946	0.1810	0.3652
	A(h_R)	0.964	0.960	0.1020	0.2347	0.970	0.953	0.1880	0.3638
	A(h_M)	0.946	0.959	0.0965	0.2253	0.961	0.941	0.1771	0.3606
n=6000	New	0.937	0.954	0.0737	0.1795	0.945	0.968	0.1371	0.3089
	A(h_S)	0.947	0.958	0.0796	0.1851	0.959	0.945	0.1462	0.2950
	A(h_R)	0.953	0.959	0.0819	0.1899	0.968	0.952	0.1513	0.2978
	A(h_M)	0.945	0.956	0.0779	0.1831	0.954	0.943	0.1435	0.2910

		p=0.9 $C(Q^L)$	$C(Q^U)$	$W(Q^L)$	$W(Q^U)$
n=1000	New	0.954	0.953	0.1762	0.3609
	A(h_S)	0.998	0.956	0.2886	0.3844
	A(h_R)	1.000	0.956	0.3218	0.3793
	A(h_M)	0.997	0.962	0.2606	0.3843
n=2000	New	0.954	0.954	0.1229	0.2550
	A(h_S)	0.995	0.958	0.1868	0.2788
	A(h_R)	0.999	0.953	0.2103	0.2765
	A(h_M)	0.989	0.954	0.1710	0.2745
n=4000	New	0.948	0.950	0.0854	0.1797
	A(h_S)	0.996	0.963	0.1210	0.1960
	A(h_R)	0.999	0.970	0.1368	0.2007
	A(h_M)	0.991	0.959	0.1122	0.1915
n=6000	New	0.940	0.964	0.0697	0.1476
	A(h_S)	0.992	0.972	0.0952	0.1578
	A(h_R)	0.998	0.976	0.1074	0.1642
	A(h_M)	0.983	0.964	0.0888	0.1546

Appendix D. Tables of simulation results for $Q_{\mathbf{TE}}(p)$

<Model 1>

		p=0.1 C(Q^L)	C(Q^U)	W(Q)	p=0.3 C(Q^L)	C(Q^U)	W(Q)	p=0.5 C(Q^L)	C(Q^U)	W(Q)
n=1000	New	0.949	0.941	4.3596	0.963	0.943	3.5685	0.957	0.949	3.3961
	$CI_{\mathrm{FP}}(h_S)$	0.901	0.884	4.2726	0.918	0.887	3.5064	0.916	0.888	3.3400
	$CI_{\mathrm{FP}}(h_R)$	0.902	0.883	4.2728	0.918	0.887	3.5065	0.916	0.889	3.3401
	$CI_{\mathrm{FP}}(h_M)$	0.894	0.882	4.2693	0.915	0.884	3.5039	0.913	0.886	3.3376
n=2000	New	0.951	0.921	4.2578	0.960	0.939	3.4905	0.947	0.948	3.3242
	$CI_{\mathrm{FP}}(h_S)$	0.898	0.878	4.2043	0.923	0.896	3.4533	0.914	0.896	3.2904
	$CI_{\mathrm{FP}}(h_R)$	0.898	0.878	4.2044	0.923	0.896	3.4534	0.914	0.897	3.2905
	$CI_{\mathrm{FP}}(h_M)$	0.897	0.876	4.2027	0.921	0.895	3.4520	0.913	0.893	3.2891
n=4000	New	0.959	0.938	4.1849	0.940	0.947	3.4357	0.939	0.957	3.2710
	$CI_{\mathrm{FP}}(h_S)$	0.904	0.908	4.1528	0.912	0.919	3.4131	0.897	0.919	3.2507
	$CI_{\mathrm{FP}}(h_R)$	0.904	0.908	4.1528	0.912	0.918	3.4132	0.897	0.919	3.2507
	$CI_{\mathrm{FP}}(h_M)$	0.903	0.907	4.1519	0.911	0.918	3.4124	0.896	0.919	3.2500
n=6000	New	0.953	0.945	4.1582	0.946	0.948	3.4125	0.942	0.951	3.2504
	$CI_{\mathrm{FP}}(h_S)$	0.911	0.909	4.1339	0.906	0.914	3.3956	0.899	0.921	3.2347
	$CI_{\mathrm{FP}}(h_R)$	0.912	0.909	4.1340	0.906	0.914	3.3956	0.899	0.921	3.2347
	$CI_{\mathrm{FP}}(h_M)$	0.910	0.908	4.1333	0.905	0.913	3.3951	0.895	0.920	3.2342

		p=0.7 C(Q^L)	C(Q^U)	W(Q)	p=0.9 C(Q^L)	C(Q^U)	W(Q)
n=1000	New	0.942	0.954	3.5621	0.953	0.971	4.3656
	$CI_{\mathrm{FP}}(h_S)$	0.900	0.901	3.5023	0.892	0.895	4.2665
	$CI_{\mathrm{FP}}(h_R)$	0.900	0.902	3.5025	0.893	0.895	4.2667
	$CI_{\mathrm{FP}}(h_M)$	0.894	0.896	3.4998	0.888	0.895	4.2634
n=2000	New	0.948	0.946	3.4875	0.931	0.957	4.2604
	$CI_{\mathrm{FP}}(h_S)$	0.902	0.893	3.4499	0.896	0.909	4.2042
	$CI_{\mathrm{FP}}(h_R)$	0.902	0.893	3.4500	0.896	0.909	4.2043
	$CI_{\mathrm{FP}}(h_M)$	0.900	0.889	3.4485	0.895	0.906	4.2026
n=4000	New	0.937	0.954	3.4333	0.938	0.956	4.1871
	$CI_{\mathrm{FP}}(h_S)$	0.910	0.917	3.4106	0.900	0.905	4.1529
	$CI_{\mathrm{FP}}(h_R)$	0.910	0.917	3.4107	0.900	0.906	4.1530
	$CI_{\mathrm{FP}}(h_M)$	0.910	0.915	3.4099	0.900	0.902	4.1520
n=6000	New	0.934	0.954	3.4114	0.926	0.955	4.1562
	$CI_{\mathrm{FP}}(h_S)$	0.895	0.917	3.3943	0.881	0.914	4.1310
	$CI_{\mathrm{FP}}(h_R)$	0.894	0.917	3.3943	0.881	0.914	4.1311
	$CI_{\mathrm{FP}}(h_M)$	0.893	0.917	3.3938	0.879	0.914	4.1304

<Model 2>

		p=0.1			p=0.3			p=0.5		
		C(Q^L)	C(Q^U)	W(Q)	C(Q^L)	C(Q^U)	W(Q)	C(Q^L)	C(Q^U)	W(Q)
n=	*New*	0.950	0.956	0.6552	0.952	0.943	0.6292	0.932	0.947	0.6754
1000	$CI_{\rm FP}$ (h_S)	0.923	0.890	0.6462	0.914	0.871	0.6194	0.892	0.885	0.6650
	$CI_{\rm FP}$ (h_R)	0.934	0.891	0.6469	0.923	0.873	0.6197	0.903	0.884	0.6654
	$CI_{\rm FP}$ (h_M)	0.912	0.887	0.6454	0.907	0.873	0.6189	0.883	0.884	0.6646
n=	*New*	0.945	0.936	0.6425	0.953	0.937	0.6171	0.936	0.949	0.6632
2000	$CI_{\rm FP}$ (h_S)	0.919	0.874	0.6368	0.926	0.882	0.6109	0.900	0.881	0.6569
	$CI_{\rm FP}$ (h_R)	0.923	0.875	0.6372	0.928	0.882	0.6111	0.903	0.881	0.6570
	$CI_{\rm FP}$ (h_M)	0.909	0.873	0.6364	0.920	0.883	0.6106	0.897	0.879	0.6566
n=	*New*	0.941	0.945	0.6345	0.943	0.949	0.6087	0.940	0.952	0.6545
4000	$CI_{\rm FP}$ (h_S)	0.913	0.896	0.6308	0.909	0.897	0.6049	0.902	0.911	0.6507
	$CI_{\rm FP}$ (h_R)	0.914	0.897	0.6310	0.911	0.897	0.6050	0.904	0.911	0.6509
	$CI_{\rm FP}$ (h_M)	0.902	0.897	0.6306	0.905	0.897	0.6048	0.898	0.912	0.6506
n=	*New*	0.949	0.944	0.6310	0.939	0.953	0.6049	0.926	0.935	0.6509
6000	$CI_{\rm FP}$ (h_S)	0.924	0.901	0.6282	0.903	0.910	0.6021	0.902	0.902	0.6482
	$CI_{\rm FP}$ (h_R)	0.932	0.901	0.6283	0.906	0.911	0.6022	0.902	0.901	0.6483
	$CI_{\rm FP}$ (h_M)	0.917	0.900	0.6281	0.898	0.909	0.6021	0.900	0.901	0.6481

		p=0.7			p=0.9		
		C(Q^L)	C(Q^U)	W(Q)	C(Q^L)	C(Q^U)	W(Q)
n=	*New*	0.886	0.945	0.7593	0.907	0.966	0.8847
1000	$CI_{\rm FP}$ (h_S)	0.810	0.874	0.7460	0.825	0.898	0.8695
	$CI_{\rm FP}$ (h_R)	0.812	0.875	0.7463	0.828	0.900	0.8702
	$CI_{\rm FP}$ (h_M)	0.809	0.871	0.7457	0.818	0.895	0.8687
n=	*New*	0.895	0.942	0.7464	0.893	0.948	0.8690
2000	$CI_{\rm FP}$ (h_S)	0.831	0.877	0.7383	0.833	0.905	0.8600
	$CI_{\rm FP}$ (h_R)	0.832	0.877	0.7385	0.839	0.908	0.8604
	$CI_{\rm FP}$ (h_M)	0.827	0.876	0.7382	0.827	0.902	0.8596
n=	*New*	0.894	0.954	0.7376	0.889	0.947	0.8589
4000	$CI_{\rm FP}$ (h_S)	0.825	0.907	0.7328	0.811	0.887	0.8531
	$CI_{\rm FP}$ (h_R)	0.827	0.908	0.7329	0.813	0.891	0.8533
	$CI_{\rm FP}$ (h_M)	0.824	0.906	0.7327	0.803	0.886	0.8529
n=	*New*	0.870	0.944	0.7339	0.874	0.942	0.8546
6000	$CI_{\rm FP}$ (h_S)	0.815	0.896	0.7303	0.810	0.896	0.8502
	$CI_{\rm FP}$ (h_R)	0.818	0.896	0.7303	0.811	0.897	0.8504
	$CI_{\rm FP}$ (h_M)	0.812	0.896	0.7302	0.807	0.893	0.8501

<Model 3>

		p=0.1 C(Q^L)	C(Q^U)	W(Q)	p=0.3 C(Q^L)	C(Q^U)	W(Q)	p=0.5 C(Q^L)	C(Q^U)	W(Q)
n= 1000	New	0.955	0.949	4.3453	0.956	0.947	3.3602	0.959	0.951	2.8402
	CI_{FP} (h_S)	0.904	0.933	4.2464	0.902	0.922	3.2972	0.907	0.924	2.7895
	CI_{FP} (h_R)	0.905	0.933	4.2465	0.902	0.922	3.2974	0.906	0.922	2.7896
	CI_{FP} (h_M)	0.899	0.928	4.2427	0.902	0.912	3.2946	0.908	0.914	2.7870
n= 2000	New	0.946	0.950	4.2333	0.962	0.954	3.2856	0.962	0.962	2.7775
	CI_{FP} (h_S)	0.900	0.910	4.1724	0.911	0.909	3.2443	0.921	0.918	2.7429
	CI_{FP} (h_R)	0.900	0.910	4.1725	0.911	0.909	3.2444	0.921	0.916	2.7429
	CI_{FP} (h_M)	0.900	0.906	4.1704	0.911	0.903	3.2429	0.920	0.910	2.7415
n= 4000	New	0.960	0.953	4.1621	0.945	0.961	3.2367	0.955	0.952	2.7314
	CI_{FP} (h_S)	0.905	0.930	4.1240	0.911	0.938	3.2118	0.913	0.930	2.7108
	CI_{FP} (h_R)	0.904	0.931	4.1240	0.911	0.938	3.2118	0.913	0.930	2.7108
	CI_{FP} (h_M)	0.904	0.926	4.1229	0.910	0.934	3.2110	0.913	0.924	2.7100
n= 6000	New	0.942	0.963	4.1301	0.941	0.959	3.2129	0.957	0.955	2.7116
	CI_{FP} (h_S)	0.911	0.920	4.1014	0.903	0.919	3.1943	0.905	0.929	2.6956
	CI_{FP} (h_R)	0.910	0.920	4.1014	0.903	0.919	3.1944	0.905	0.929	2.6956
	CI_{FP} (h_M)	0.910	0.917	4.1006	0.903	0.918	3.1938	0.905	0.926	2.6951

		p=0.7 C(Q^L)	C(Q^U)	W(Q)	p=0.9 C(Q^L)	C(Q^U)	W(Q)
n= 1000	New	0.952	0.964	2.5954	0.946	0.968	2.7691
	CI_{FP} (h_S)	0.902	0.931	2.5488	0.898	0.946	2.7241
	CI_{FP} (h_R)	0.902	0.931	2.5489	0.898	0.947	2.7243
	CI_{FP} (h_M)	0.903	0.922	2.5459	0.896	0.933	2.7199
n= 2000	New	0.957	0.955	2.5358	0.960	0.964	2.7090
	CI_{FP} (h_S)	0.916	0.927	2.5070	0.924	0.935	2.6802
	CI_{FP} (h_R)	0.916	0.925	2.5070	0.924	0.937	2.6802
	CI_{FP} (h_M)	0.915	0.910	2.5054	0.923	0.930	2.6779
n= 4000	New	0.949	0.959	2.4948	0.941	0.970	2.6647
	CI_{FP} (h_S)	0.918	0.933	2.4764	0.915	0.945	2.6442
	CI_{FP} (h_R)	0.918	0.934	2.4764	0.915	0.945	2.6443
	CI_{FP} (h_M)	0.918	0.930	2.4756	0.915	0.943	2.6431
n= 6000	New	0.955	0.956	2.4754	0.942	0.959	2.6451
	CI_{FP} (h_S)	0.911	0.923	2.4620	0.906	0.938	2.6297
	CI_{FP} (h_R)	0.911	0.923	2.4620	0.906	0.938	2.6297
	CI_{FP} (h_M)	0.911	0.918	2.4614	0.905	0.934	2.6289

<Model 4>

		p=0.1			p=0.3			p=0.5		
		C(Q^L)	C(Q^U)	W(Q)	C(Q^L)	C(Q^U)	W(Q)	C(Q^L)	C(Q^U)	W(Q)
n=	New	0.973	0.958	2.4848	0.959	0.956	2.2337	0.947	0.947	2.2772
1000	CI_{FP} (h_S)	0.951	0.925	2.4551	0.938	0.920	2.2131	0.939	0.930	2.2572
	CI_{FP} (h_R)	0.966	0.933	2.4664	0.954	0.933	2.2228	0.951	0.931	2.2684
	CI_{FP} (h_M)	0.946	0.919	2.4492	0.931	0.911	2.2081	0.929	0.918	2.2515
n=	New	0.956	0.949	2.4263	0.958	0.956	2.1879	0.951	0.953	2.2251
2000	CI_{FP} (h_S)	0.928	0.915	2.4073	0.936	0.928	2.1736	0.933	0.926	2.2110
	CI_{FP} (h_R)	0.942	0.930	2.4138	0.945	0.938	2.1792	0.946	0.927	2.2174
	CI_{FP} (h_M)	0.925	0.908	2.4041	0.932	0.917	2.1709	0.932	0.923	2.2080
n=	New	0.960	0.949	2.3908	0.956	0.949	2.1557	0.963	0.943	2.1889
4000	CI_{FP} (h_S)	0.934	0.925	2.3784	0.930	0.928	2.1467	0.947	0.918	2.1802
	CI_{FP} (h_R)	0.945	0.932	2.3820	0.949	0.934	2.1498	0.957	0.930	2.1838
	CI_{FP} (h_M)	0.932	0.921	2.3767	0.927	0.924	2.1453	0.942	0.912	2.1786
n=	New	0.951	0.943	2.3736	0.942	0.949	2.1410	0.941	0.946	2.1729
6000	CI_{FP} (h_S)	0.928	0.918	2.3639	0.928	0.933	2.1342	0.925	0.925	2.1660
	CI_{FP} (h_R)	0.935	0.925	2.3665	0.936	0.939	2.1365	0.934	0.936	2.1686
	CI_{FP} (h_M)	0.926	0.914	2.3627	0.925	0.930	2.1332	0.922	0.925	2.1648

		p=0.7			p=0.9		
		C(Q^L)	C(Q^U)	W(Q)	C(Q^L)	C(Q^U)	W(Q)
n=	New	0.952	0.971	3.0731	0.960	0.965	2.2546
1000	CI_{FP} (h_S)	0.979	0.936	2.8786	0.990	0.985	2.2873
	CI_{FP} (h_R)	0.972	0.935	2.8548	0.999	0.989	2.3151
	CI_{FP} (h_M)	0.978	0.918	2.8720	0.981	0.968	2.2689
n=	New	0.946	0.959	2.8095	0.941	0.946	2.2110
2000	CI_{FP} (h_S)	0.953	0.933	2.7434	0.982	0.964	2.2280
	CI_{FP} (h_R)	0.964	0.933	2.7458	0.995	0.980	2.2469
	CI_{FP} (h_M)	0.946	0.926	2.7366	0.971	0.954	2.2177
n=	New	0.959	0.949	2.6824	0.959	0.960	2.1821
4000	CI_{FP} (h_S)	0.963	0.918	2.6529	0.979	0.970	2.1902
	CI_{FP} (h_R)	0.972	0.921	2.6597	0.993	0.978	2.2018
	CI_{FP} (h_M)	0.960	0.912	2.6490	0.968	0.961	2.1844
n=	New	0.940	0.951	2.6416	0.955	0.962	2.1670
6000	CI_{FP} (h_S)	0.938	0.931	2.6209	0.977	0.971	2.1726
	CI_{FP} (h_R)	0.953	0.934	2.6268	0.994	0.983	2.1815
	CI_{FP} (h_M)	0.934	0.930	2.6182	0.970	0.956	2.1684

<Model 5>

		p=0.1 C(Q^L)	C(Q^U)	W(Q)	p=0.3 C(Q^L)	C(Q^U)	W(Q)	p=0.5 C(Q^L)	C(Q^U)	W(Q)
n=	*New*	0.982	0.967	1.5649	0.960	0.940	0.5396	0.953	0.952	0.4245
1000	CI_{FP} (h_S)	0.920	0.932	1.4107	0.959	0.989	0.5340	0.994	0.991	0.4389
	CI_{FP} (h_R)	0.926	0.935	1.4366	0.911	0.921	0.5199	0.937	0.930	0.4192
	CI_{FP} (h_M)	0.921	0.936	1.4397	0.910	0.909	0.5194	0.936	0.924	0.4182
n=	*New*	0.980	0.966	1.4295	0.965	0.956	0.5130	0.957	0.956	0.4127
2000	CI_{FP} (h_S)	0.916	0.937	1.3447	0.956	0.995	0.5109	0.982	0.981	0.4209
	CI_{FP} (h_R)	0.929	0.937	1.3586	0.925	0.930	0.5027	0.933	0.942	0.4093
	CI_{FP} (h_M)	0.924	0.937	1.3591	0.922	0.924	0.5024	0.923	0.931	0.4087
n=	*New*	0.974	0.961	1.3423	0.960	0.952	0.4979	0.958	0.949	0.4049
4000	CI_{FP} (h_S)	0.924	0.936	1.2944	0.952	0.983	0.4968	0.989	0.983	0.4096
	CI_{FP} (h_R)	0.930	0.933	1.3014	0.933	0.927	0.4921	0.937	0.933	0.4029
	CI_{FP} (h_M)	0.929	0.933	1.3016	0.932	0.921	0.4920	0.934	0.926	0.4025
n=	*New*	0.966	0.952	1.3044	0.945	0.946	0.4912	0.947	0.962	0.4015
6000	CI_{FP} (h_S)	0.905	0.928	1.2716	0.942	0.983	0.4903	0.974	0.984	0.4048
	CI_{FP} (h_R)	0.913	0.929	1.2761	0.915	0.926	0.4870	0.930	0.933	0.3999
	CI_{FP} (h_M)	0.914	0.928	1.2762	0.913	0.923	0.4868	0.923	0.928	0.3997

		p=0.7 C(Q^L)	C(Q^U)	W(Q)	p=0.9 C(Q^L)	C(Q^U)	W(Q)
n=	*New*	0.957	0.963	0.5386	0.970	0.980	1.5684
1000	CI_{FP} (h_S)	0.995	0.953	0.5320	0.928	0.915	1.4130
	CI_{FP} (h_R)	0.930	0.906	0.5178	0.932	0.922	1.4650
	CI_{FP} (h_M)	0.926	0.903	0.5172	0.931	0.924	1.4927
n=	*New*	0.950	0.949	0.5135	0.965	0.974	1.4299
2000	CI_{FP} (h_S)	0.991	0.933	0.5104	0.934	0.910	1.3443
	CI_{FP} (h_R)	0.926	0.901	0.5022	0.935	0.919	1.3625
	CI_{FP} (h_M)	0.922	0.900	0.5019	0.930	0.918	1.3646
n=	*New*	0.947	0.950	0.4978	0.957	0.972	1.3420
4000	CI_{FP} (h_S)	0.984	0.942	0.4963	0.935	0.928	1.2935
	CI_{FP} (h_R)	0.923	0.917	0.4916	0.934	0.931	1.3012
	CI_{FP} (h_M)	0.912	0.916	0.4915	0.934	0.930	1.3017
n=	*New*	0.939	0.962	0.4914	0.960	0.963	1.3030
6000	CI_{FP} (h_S)	0.983	0.948	0.4904	0.928	0.917	1.2706
	CI_{FP} (h_R)	0.923	0.930	0.4871	0.929	0.927	1.2752
	CI_{FP} (h_M)	0.917	0.929	0.4870	0.929	0.927	1.2755

<Model 6>

		p=0.1			p=0.3			p=0.5		
		C(Q^L)	C(Q^U)	W(Q)	C(Q^L)	C(Q^U)	W(Q)	C(Q^L)	C(Q^U)	W(Q)
n=	*New*	0.955	0.945	2.8916	0.953	0.956	2.7884	0.948	0.954	3.2540
1000	CI_{FP} (h_S)	0.944	0.913	2.8595	0.946	0.923	2.7568	0.941	0.924	3.2099
	CI_{FP} (h_R)	0.954	0.914	2.8681	0.950	0.926	2.7655	0.952	0.924	3.2192
	CI_{FP} (h_M)	0.935	0.910	2.8527	0.938	0.917	2.7504	0.934	0.917	3.2022
n=	*New*	0.948	0.946	2.8246	0.951	0.943	2.7200	0.941	0.945	3.1611
2000	CI_{FP} (h_S)	0.931	0.917	2.8032	0.944	0.911	2.6988	0.938	0.917	3.1355
	CI_{FP} (h_R)	0.938	0.919	2.8080	0.948	0.913	2.7038	0.948	0.919	3.1419
	CI_{FP} (h_M)	0.927	0.916	2.7995	0.931	0.906	2.6955	0.934	0.913	3.1314
n=	*New*	0.956	0.946	2.7804	0.946	0.956	2.6748	0.934	0.963	3.1071
4000	CI_{FP} (h_S)	0.937	0.911	2.7663	0.935	0.931	2.6614	0.925	0.946	3.0903
	CI_{FP} (h_R)	0.946	0.913	2.7692	0.940	0.934	2.6645	0.936	0.949	3.0943
	CI_{FP} (h_M)	0.931	0.905	2.7643	0.930	0.930	2.6597	0.914	0.945	3.0880
n=	*New*	0.941	0.959	2.7584	0.937	0.951	2.6533	0.920	0.958	3.0813
6000	CI_{FP} (h_S)	0.913	0.930	2.7478	0.923	0.926	2.6433	0.914	0.938	3.0687
	CI_{FP} (h_R)	0.923	0.931	2.7499	0.932	0.928	2.6455	0.919	0.941	3.0717
	CI_{FP} (h_M)	0.907	0.928	2.7464	0.914	0.922	2.6420	0.911	0.936	3.0672

		p=0.7			p=0.9		
		C(Q^L)	C(Q^U)	W(Q)	C(Q^L)	C(Q^U)	W(Q)
n=	*New*	0.937	0.956	4.7842	0.963	0.962	3.5672
1000	CI_{FP} (h_S)	0.977	0.889	4.5115	0.991	0.936	3.5658
	CI_{FP} (h_R)	0.975	0.898	4.4986	0.998	0.931	3.5780
	CI_{FP} (h_M)	0.969	0.887	4.5142	0.987	0.936	3.5541
n=	*New*	0.954	0.957	4.5973	0.964	0.955	3.4898
2000	CI_{FP} (h_S)	0.963	0.898	4.4270	0.989	0.929	3.4886
	CI_{FP} (h_R)	0.974	0.900	4.4241	0.999	0.925	3.4979
	CI_{FP} (h_M)	0.960	0.889	4.4238	0.980	0.927	3.4801
n=	*New*	0.948	0.961	4.4279	0.958	0.947	3.4402
4000	CI_{FP} (h_S)	0.949	0.913	4.3609	0.980	0.929	3.4389
	CI_{FP} (h_R)	0.959	0.923	4.3630	0.995	0.936	3.4478
	CI_{FP} (h_M)	0.946	0.910	4.3574	0.973	0.926	3.4333
n=	*New*	0.949	0.953	4.3704	0.956	0.957	3.4160
6000	CI_{FP} (h_S)	0.949	0.908	4.3316	0.979	0.954	3.4152
	CI_{FP} (h_R)	0.955	0.915	4.3348	0.991	0.956	3.4229
	CI_{FP} (h_M)	0.946	0.906	4.3288	0.963	0.948	3.4112

CHAPTER V

HETEROGENEOUS EFFECTS OF CLASS SIZE REDUCTION: RE-VISITING PROJECT STAR

Introduction

In previous chapters, I presented econometric theoretical developments on partial identification and the statistical inference on identifying the distribution of treatment effects and their quantiles in the context of a randomized experiment. In this chapter, I will apply those to the field data. The randomized experiment I consider in this chapter is Project STAR. Project STAR, the acronym of The Student/Teacher Achievement Ratio, is a large scale, longitudinal, randomized experiment conducted by the Tennessee State Department of Education during 1985 - 1989 in order to see if the class size reduction improves students' academic attainment.

Based on their findings, Glass and Smith suggested 1:15 is the optimal students-teacher ratio and reducing the class size to that ratio would improve students' achievement (Glass and Smith 1978, 1979). If this is true, the class sizes at that time were 33%-133% larger than the optimal size. However, class size reduction (CSR) incurs costs. CSR means more teachers, educational instruments, buildings, and schools. Lacking conclusive evidence, the Tennessee state government had decided to conduct a large scale experiment on the effect of CSR in elementary schools prior to implementation, called Project STAR.

This chapter uses the data from Project STAR. Particularly, I follow Ding and Lehrer (2005) in focusing on the achievement in kindergarten because the randomness of the experiment after the completion of kindergarten seemed contaminated by intent to treat, selective attrition, and new participants in each grade. In terms of randomness, the kindergarten achievement is the most intact.[1] See Ding and Lehrer (2004).

A large existing literature has investigated the effects of CSR using the same data. With few exceptions, these papers sough to identify the average treatment effects (ATE) of the CSR. A key exception in the literature that used Project STAR data is Ding and Lehrer (2005). Instead of identifying the ATE of CSR, they investigated the heterogeneity of treatment effects in CSR by using quantile regressions and found strong evidences that effects of CSR are heterogeneous on the ability of students. Based on the results, they concluded, "higher ability students gain the most from CSR while many low ability students do not benefit from these reductions."

Essentially, they estimated the 'quantile treatment effects' (QTE). As pointed out in Chapter III and IV, the QTE agrees the quantiles of treatment effects ($\mathrm{Q_{TE}}$) only under specific assumptions on the dependent structure between treated and controlled outcomes. Therefore, their conclusion based on the estimation of QTE may not hold in general. In this chapter, I will adopt a partial identification approach in order to identify the distribution of treatment effects of CSR and also provide a new way to looking at the heterogeneous treatment effects.

The rest of the chapter is organized as follows. In Section 2, I will briefly describe about Project STAR, its historical backgrounds, and its findings including results of recent research. In Section 3, I will summarize some of the key findings from the previous chapters that I will use in this chapter and add some additional concepts and a theorem which are important in analyzing the CSR. The empirical findings are discussed in Section 4 and

[1] Even kindergarten achievement needs cautious interpretation. As Ding and Lehrer (2005) pointed out, Kindergarten was not mandatory in Tennessee. Potentially, different group of people may have had different level of accessibility to Kindergarten. In other words, Project STAR experiment could not control for selection at the very initial level. However, the pursuit of the issue is beyond the scope of this paper.

Section 5 concludes.

Project STAR

Brief Historical Background

The effect of the CSR has been debated for decades but the conclusion was controversial. Some surveys suggested theoretical channels through which smaller classes helped students attain higher scores.

Hallinan and Sorensen (1985) reported that teachers' morale and job satisfaction are higher in small classes and teachers reported that students had better attitudes and motivations. Filby, Cahen, McCutsheon, and Kyle (1980) found teachers were more able to help students when they needed it in smaller classes. In the survey, teachers responded that their work load became lighter, which enabled them to make the classroom climate more positive.

However, empirical research did not provide conclusive evidences. In their Meta-Analysis with 77 existing studies, Glass, Cahen, and Smith (1978) asserted that they found a trend that the students' achievement decreases as class size increased and they claimed the greatest gains occurred when student-teacher ratio was 1:15 or below. On the contrary, Robinson and Wittebols (1986) found only 35 studies out of the 85 they considered to be relevant reported small classes were better, 18 supported larger classes, and the rest 32 did not support either.

Prior to the launch of Project STAR, Whittington, Bain, and Achilles (1985) investigated the effect of CSR from 1:25 to 1:15 by doing a small scale experimental study with first grade students in the Metro Nashville School District. They reported the students in classes of 15 students performed better than those in classes of 25 in reading and math. On the other hand, Dennis (1986) could not observe any difference between the treated group and control groups in the following year. Bourke (1986) found the class size itself did not affect students' attainment directly. It was, he claimed, teachers' practices that enhanced student achievement. Moreover, teachers do not change their teaching practices when class size is reduced. (Robinson 1990).

There continue to be debates on the effectiveness of CSR. For example, Hanushec (1998) could find "little reason to believe that smaller class sizes systematically yield higher student achievement", while Krueger (2002) found exactly the opposite and said "when studies are assigned weights in proportion to the 'impact factor'... class size is systematically related to achievement."

Because the CSR was costly, and the results of proceeding researches were not conclusive, the Tennessee State Government decided to conduct a well-designed randomized experiment to investigate whether or not the CSR would be effective before implementing the CSR. In May, 1985, the Tennessee Legislature passed House Bill (HB) 544, which authorized and funded an experimental study on the effect of CSR, which was Project STAR.

The project was conducted by a consortium of persons from Tennessee State Department of Education, Memphis State University, Tennessee State University, University of Tennessee at Knoxville, Vanderbilt University, representatives from the State Board of Education and the State Superintendents' Association. Only a few months after the pass of the legislation, the consortium was able to set up major parts of the project and to implement it from the fall semester of 1985-1986 schooling year, which continued up until 1988-1989 schooling year.

Design of Project STAR

Tennessee had been regulating the student/teacher ratio even before Project STAR.[2] By the time Project STAR started, the ratio could not exceed 1:25. The average of the ratios was 1:22-24. The legislation regulated the ratio in small classes to be between 1:13 and 1:17. So the main question that Project STAR should answer was whether 1:13-17 would be better than 1:22-24 for students' academic achievements.

The consortium decided to divide the class sizes/environment into three categories; small class (1:13-17), regular class (1:22-25), and regular class with teacher aide (1:22-25). The consortium chose 79 schools from four different areas; inner-city (schools located in metropolitan areas plus those that had more than half of their students on free of reduced cost lunch), suburban (schools located in the outlying areas of metropolitan cities), urban (schools located in a town of over 2,500 and serving primarily an urban population), and rural (all other schools).

The project schools were chosen out of 180 volunteers from 141 school systems all over the state. Because the consortium designed the project to make the 'within-school' comparison available as well, each school had to have certain number of students so that it had at least one class of each type. After an investigation, the consortium chose 79 schools as the participants for the 1985-1986 schooling year. Initially, a school should plan to remain in the project for the whole years but 1-3 schools left. The initial objective was to have about 100 classes of each type. In the first year, there were 128 small classes, 101 regular classes, and 99 regular classes with teacher aide. Each participating school had to agree to assign students and teachers randomly in three types of classes and not to make any significant changes in their provision of education other than class size. Roughly 6,000

[2]This subsection is a summary of the technical report of the STAR project (Word et. al. 1990b), which I will refer as *Technical Report*.

students participated in the project every year.

Table 22 is the number of participating schools.

Table 22. Project Star Schools

	Kindergarten	Grade 1	Grade 2	Grade 3
Inner City	17	15	15	15
Suburban	16	15	15	15
Rural	38	38	38	38
Urban	8	8	7	7
Total	79	76	75	75

* Technical Report, p6.

To identify whether or not the Hawthorne Effect[3] existed, the consortium chose comparison group from non-participant schools and compared their achievements with the counterparts of participants. They found the Hawthorne Effect was not a significant factor.

The kindergarten student academic achievements were measured by Stanford Early School Achievement Test II (SESAT II) in Math, Sounds and Letters, Words and Sentences, and Reading. Higher graders used the Stanford Achievement Tests (SAT), the State of Tennessee's criterian-referenced Basic Skills First (BSF) tests. In this chapter, I will use the data of kindergarten students on math and reading only.

Effects of CSR

After the project was over, Word reported the followings (Word et. al. (1990b), pp.17-19).

1. Small-class advantages are present in all locations and all grades. Students in small classes showed higher performance than those in regular classes or regular classes with

[3]Originally, the Hawthorne Effect was coined by Henry A. Landsberger (1955) in order to indicate"a short-term improvement caused by observing worker performance." However, the definition is broadened in various scholarly discplines. Generally, the Hawthorn Effect refers to the changes in behaviors of people in response to the attention they receive (see Jex 2000).

In the context of the STAR project, this effect means systematic differences between the participant schools and non-participant schools even though all other things were equal between them. If this effect exists, the differences between S and R, for example, may also be contaminated by this effect.

teacher aide.

2. Small-class effects diminish after first grade but are significant at the end of third grade.

3. Teacher aides were less effective than CSR in enhancing student performance at each grade level.

4. The effects in Math and reading are similar.

5. Small classes help low socioeconomic students as much as they helped high socioeconomic students. In reading, low socioeconomic students appeared to benefit more whereas in math the high socioeconomic student did.[4]

A follow-up study on the mid term effects of CSR found a significantly larger proportion of the small class students than regular or regular with aids class students had passed the Tennessee Competency Examination (TCE) requirement at eighth grade. (Pate-Bain et. al. 1997) Another follow-up study showed similar results. Students in small classes were more likely to take ACT or SAT exams and the difference in proportions of students who took a college entrance exam out of Project STAR participants differed across races. The difference was larger for black students, which indicates CSR benefited black students more in the long run. In addition, the average scores of small class students were significantly higher than that of large class students (Krueger and Whitmore 2001).

Including the project reports, almost all of existing literature focuses on the average effect of treatment (CSR) gains with the consideration of observable heterogeneities such as sex, age, race, and the socioeconomic status. One notable exception is Ding and Lehrer

[4]Students' socioeconomic status is measured by a dummy variable indicating whether or not student joined a free or reduced price lunch program. If they joined, they were considered to be poor or of low socioeconomic status.

(2005). They estimated the following equation with the kindergarten data:

$$Y_{ij} = \alpha' X_{ij} + \delta' CS_{ij} + v_j + \varepsilon_{ij}$$

where Y_{ij} is the level of achievement for child i in school j, X_{ij} a vector of student and teacher characteristics, CS_{ij} the actual number of students in the class where child i belonged to, v_j a school fixed effect, and ε_{ij} the random and idiosyncratic unobservable factors. They used quantile regression to estimate δ for different quantiles of Y_{ij}. Each estimate of δ, for both math and reading, over different quantiles have negative signs, indicating that the student outcomes decreases as class size increases.

More interesting are the magnitudes of δ's. They increase generally as quantile level increases. For example, δ for the 0.05 quantile is about -0.15 whereas that for the 0.95 quantile is about -1.85 in math. In reading, δ for the 0.05 quantile is about 0 but that for the 0.95 quantile is about -1.3. Ding and Lehrer interpreted this as an evidence that higher ability students gain more from the CSR. Their interpretation, though, is only valid when we consider QTE as if they are $\mathrm{Q_{TE}}$. They implicitly assumed so-called rank preservation property in the sense that those δ's measure the effects of CSR between the same quantiles.

In this chapter, I will re-examine their findings by adopting the partial identification framework discussed in the previous three chapters in regard with the $\mathrm{Q_{TE}}$.

Partial Identification of the Parameters of Interest

Quantiles of Treatment Effects

As in the previous chapters, I will use Y_1 and Y_0 to denote the (potential) outcomes from treatment and control respectively. The distributions of Y_1 and Y_0 are F_1 and F_0

respectively. The outcomes are standardized test scores in math and reading. I considered small class (S hereafter) and regular class with aide (A) as two treatments and regular class (R) as the control as if S and A were two potential policies being considered to alter R, the status quo. The treatment effects, Δ, is the outcome gap between S and R or between S and A depending on the situation.

The first parameter of interest is $Q(p) = \arg\inf_{\delta}\{F_{\Delta}(\delta) \geq p\}$, the quantile function of Δ. As discussed in Chapter III, $Q(p)$ is partially identified with the identification region $\left[Q^L(p), Q^U(p)\right]$ for any $p \in (0,1)$, where

$$\begin{aligned} Q^U(p) &= \inf_{u \in (p,1)} [F_1^{-1}(u) - F_0^{-1}(u-p)], \\ Q^L(p) &= \sup_{u \in (0,p)} [F_1^{-1}(u) - F_0^{-1}(1+u-p)]. \end{aligned}$$

I proposed the nonparametric estimators of $Q^L(p)$ and $Q^U(p)$, $Q_n^L(p)$ and $Q_n^U(p)$ respectively,

$$Q_n^L(p) = \sup_{u \in \left[\frac{1}{\max\{n_1, n_0\}}, p\right]} \{F_{1n}^{-1}(u) - F_{0n}^{-1}(1+u-p)\}, \tag{V.1}$$

$$Q_n^U(p) = \inf_{u \in \left[p, 1-\frac{1}{\max\{n_1, n_0\}}\right]} \{F_{1n}^{-1}(u) - F_{0n}^{-1}(u-p)\}, \tag{V.2}$$

and their joint asymptotic distribution under (A1)-(A4), which is;

$$\sqrt{n_1}\begin{pmatrix} Q_n^L(p) - Q^L(p) \\ Q_n^U(p) - Q^U(p) \end{pmatrix} \Rightarrow N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_L^2 & \sigma_{LU} \\ \sigma_{LU} & \sigma_U^2 \end{pmatrix}\right)$$

when iid samples of Y_1 and Y_0 (of size n_1 and n_0 respectively) are available.

Of the methods to construct confidence intervals, I will use CI_{FP} and the extension of 'new approach' (CI_{NEW}) in this chapter. For each $p \in (0,1)$ and $\alpha < \frac{1}{2}$,

$$\begin{aligned} CI_{\text{FP}} &= \left\{ x : T_n(p) \leq cv_{1-\alpha}\left(\frac{\sqrt{n_1}\left(Q_n^U(p) - Q_n^L(p)\right)}{\max\left\{\hat{\sigma}_L, \hat{\sigma}_U\right\}}, \hat{\rho} \right) \right\}, \\ CI_{\text{NEW}} &= \left[Y_{1\left(\left[n_1 \hat{u}_{1nL}^A\right]\right)} - Y_{0\left(\left[n_0 \hat{u}_{0nL}^A\right]\right)}, Y_{1\left(\left[n_1 \hat{u}_{1nL}^B\right]\right)} - Y_{0\left(\left[n_0 \hat{u}_{0nL}^B\right]\right)} \right] \end{aligned}$$

Refer to the previous chapters for the definitions of $\hat{\rho}$, $\hat{\sigma}_L$, $\hat{\sigma}_U$, $\hat{u}_{1nL}^A$, $\hat{u}_{0nL}^A$, $\hat{u}_{1nL}^B$, and $\hat{u}_{0nL}^B$. The Δ and $\hat{\Delta}$ in Chapter I mean $Q^U(p) - Q^L(p)$ and $Q_n^U(p) - Q_n^L(p)$ in regard of current chapter's discussion.

$Q(p)$ is the outcome gain that p-th quantile person benefits. Therefore $Q^L(p)$ and $Q^U(p)$ are the minimum and maximum amounts of outcome gain of p-th quantile person. If $Q^U(p) \leq 0$ for a p then at least $100 * p$ % of population will lose or not gain at the best from the treatment. On the other hand, if $Q^L(p) \geq 0$ for a p then it means $100 * (1-p)$ % of population will gain nonnegative amount from the treatment.

More generally, let $p_U = \sup_{p\in(0,1)} Q^U(p) \leq 0$ and $p_L = \inf_{p\in(0,1)} Q^L(p) \geq 0$. Let p_+ be the proportion of population who do not get hurt and p_- the proportion who do not benefit from the treatment.

Lemma 8 *p_+ and p_- are partially identified. Their sharp identification regions are;*

$$\begin{aligned} p_+ &\in \left[1 - p_L, 1 - p_U\right], \\ p_- &\in \left[p_U, p_L\right]. \end{aligned}$$

The concept of p_+ and p_- can serve as a criterion to choose a policy among many candidates. Imagine two policies S and A are being considered to alter a status quo R. Let $Q_{R\to S}$ be the quantile functions of outcome gains from the policy change of R to S and $Q_{R\to A}$ the quantile function of outcome gain from R to A. If we can observe the $Q_{R\to S}$ and $Q_{R\to A}$ then we can adopt stochastic dominance to evaluate whether $R \to S$ change will be better than $R \to A$ change or vice versa. However, when $Q_{R\to S}$ and $Q_{R\to A}$ are only partially identified as is in the chapter, we cannot have a dominance ranking between two changes.

However, we may use p_L and p_U as our criteria. If policy makers focus on the minimizing the number of people who get hurt by policy changes, they may want to choose the policy that has the least p_U because larger p_U implies larger fraction of population will get hurt for sure by the change. If, instead, policy makers focus on the number of people who benefit from the policy change, they may want to choose a change that has the smallest p_L because by doing so they can maximize the minimum proportion of population who gain from the treatment.

Another use of $\left[Q^L(p), Q^U(p)\right]$ is that the region can provide a test for the 'common treatment effects' or 'homogeneous treatment effects'. If the treatments are constant over all p then $Q(p)$ is constant for all $p \in (0,1)$ and $\sup Q^L(p) \geq \inf Q^U(p)$. Graphically, Figure 11(A) allows for $Q(p)$ be a constant while Figure 11(B) does not. Therefore, if $\sup Q^L(p) < \inf Q^U(p)$ then we can reject the homogeneous treatment effects hypothesis.

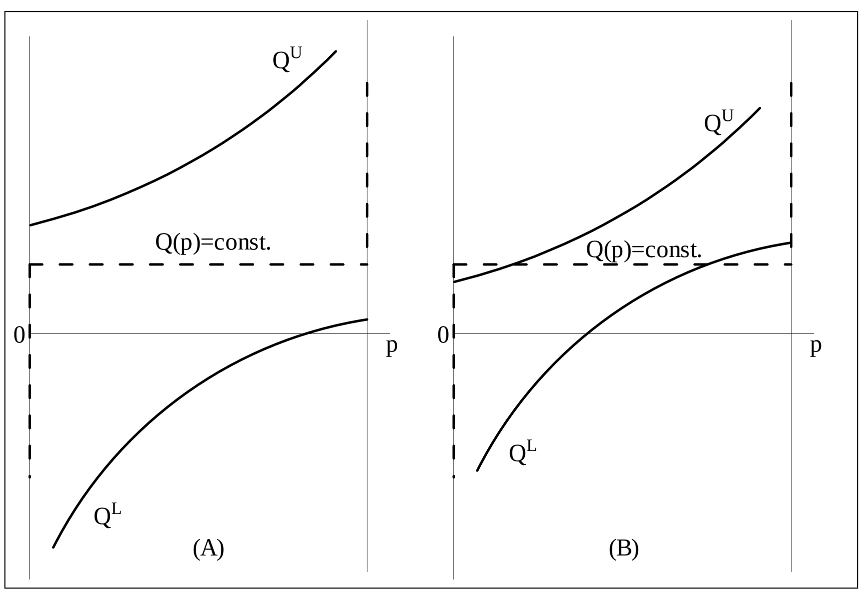


Figure 11. Concept of Testing for 'Homogeneous Treatment Effects'

In practice, we have Q_n^L and Q_n^U instead of Q^L and Q^U. So, we can use the following steps to see if the homogeneous treatment hypothesis is consistent with data.

1) Construct confidence sets for $Q(p)$ for $p \in (0,1)$ at a given significant level of α.

2) Find the maximum of the lower bounds and the minimum of the upper bounds of confidence intervals for all p at a given α.

3) Conclude the homogeneous treatment hypothesis is not consistent with the data available if the maximum of the lower bounds is larger than the minimum of the upper bounds.

Bounds on the Conditional Distribution of Treatment Effects upon Pre-Treatment Outcomes

In some cases, researchers may be interested in the changes themselves of outcomes rather than the output gains due to a treatment. For example, researchers may want to know how a policy changes lowest quintile people's income instead of net outcome gains. The knowledge on the distribution of outcome gains conditional upon the pre-treatment outcomes can provide insightful information on that. Let $\Delta_i = Y_{1i} - Y_{0i}$ as in Chapter II and III where $Y_{1i} \sim_{iid} F_1$ and $Y_{0i} \sim_{iid} F_0$ for all i. Then $\Pr[\Delta_i \geq 0 | Y_{0i} \in B]$ tells us the probability that an individual obtains no loss from treatment when his/her pre-treatment outcome is in B. If we can identify this conditional probability, it can help us to better understand the heterogeneity in the treatment effects. The following definitions and lemmas formulate basic ideas.[5]

Definition 3 *If Δ is a constant, the treatment effects are homogeneous. If Δ is independent of Y_0, the treatment effects are weakly homogeneous. If treatment effects are neither homogeneous nor weakly homogeneous, they are heterogeneous.*

[5]The concepts here are defined without any covariate. But the basic ideas can be extendable trivially to the covariates by considering $\Delta|X$ instead of Δ. For instance, the concept of weak homogeneity can be extended as the independence of $\Delta|X$ of Y_0. In other words, the heterogeneity being considered here is the heterogeneity on unobservables.

The weak homogeneity is an extension of homogeneity of treatment effects. Under the assumption of weak homogeneity, F_Δ is point-identified. Heckman, Smith, and Clements (1997) showed

$$F_\Delta(\delta) = \frac{1}{2} + \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{1}{it}\left(e^{it\delta}\frac{E\left[e^{-itY_1}\right]}{E\left[e^{-itY_0}\right]} - e^{-it\delta}\frac{E\left[e^{itY_1}\right]}{E\left[e^{itY_0}\right]}\right)dt$$

when $\Delta \perp Y_0|D = 1$.

If treatment effects are weakly homogeneous, then $E[\Delta|Y_0] = E[\Delta]$. The homogeneity of treatment effects implies the weak homogeneity but not vice versa. From the definition, the following two lemmas hold. Let us define $\Pr[\Delta \leq \delta|Y_0 \in \varnothing] = 0$ for all δ.

Lemma 9 *If the treatment effects are weakly homogeneous,* $\Pr[\Delta \leq \delta|Y_0 \in B]$ *is not a function of B.*

Lemma 10 *The treatment effects are homogeneous if and only if* $\Pr[\Delta \leq \delta|Y_0 \in B]$ *is either 0 or 1 for all δ and B.*

The use of the conditional probability of Δ on Y_0 allows us to develop the concept of the conditional heterogeneity of Δ. First, define two functions

$$\Lambda_\delta(B) = \Pr[\Delta > \delta|Y_0 \in B] \text{ and } \Psi_\delta(B) = \Pr[\Delta > \delta|Y_0 \notin B].$$

I will focus on $\Lambda_\delta(y) \equiv \Pr[\Delta \geq \delta|Y_0 \leq y]$ and $\Psi_\delta(y) \equiv \Pr[\Delta \geq \delta|Y_0 \geq y]$.[6] In words, $\Lambda_\delta(y)$ is the probability that an individual benefits by no less than δ when his/her potential 'controlled outcome' or 'pre-treatment outcome' is at most y. Or it is the proportion of 'poor' people (in the sense that they earn no more than y in pre-treatment state) who will gain by δ or more if the treatment is implemented.[7] On the other hand, $\Psi_\delta(y)$ is the probability that the gains from treatment is at least δ given that the pre-treatment outcome is no less than y. So, in a sense, it tells us the proportion of 'rich' people (those

[6] I assume Y_0 is continuous here. Further, I will assume Y_1 is also continuous.

[7] Interpreting $\Lambda_\delta(y)$ in this way, I have in mind a situation that a society tests a policy (treatment) in a small-sized randomized experiment situation prior to implementation. To highlight this interpretation, I will use 'pre-treatment outcome' from this point forward.

who are earning no less than y in pre-treatment state)' whose gains from treatment are no less than δ.

While the concept of $Q(p)$ or $F_{\Delta}(\delta)$ focuses on 'how many people will gain a certain amount of outcome difference when a policy is implemented', $\Lambda_{\delta}(B)$ and $\Psi_{\delta}(B)$ focus more on 'how people's outcome will change after the policy implementation when their current outcomes are in a specific range.'

Definition 4 *If $\Lambda_{\delta}(y) = \Pr[\Delta \geq \delta | Y_0 \leq y]$ is not a function of y, the treatment effects above δ are conditionally homogeneous at the lower margin. If $\Psi_{\delta}(y) = \Pr[\Delta \geq \delta | Y_0 \geq y]$ is not a function of y, the treatment effects above δ are conditionally homogeneous at the upper margin. If they are not conditionally homogeneous on respective margins, they are conditionally heterogeneous at the respective margins.*

The two functions measure the conditional heterogeneity in different ways. Because $\lim_{y\to\infty} \Lambda_{\delta}(y) = \Pr[\Delta \geq \delta]$, $\Lambda_{\delta}(y)$ measured at a large value of y does not provide meaningful information of the conditional heterogeneity on pre-treatment outcomes. It mainly provides information on conditional heterogeneity at the lower quantiles of Y_0. $\Psi_{\delta}(y)$ is the opposite. It contains significant information of the conditional heterogeneity at the upper quantiles of Y_0. The information that $\Lambda_{\delta}(y)$ and $\Psi_{\delta}(y)$ carry is meaningful. To see this, assume $\Lambda_{\delta}(y)$ and $\Psi_{\delta}(y)$ are decreasing in y. That means the 'poorer' is a person in pre-treatment state, the higher is the change of benefitting by δ or more from the treatment. With the slight abuse of the concept of being progressive or regressive, I will define the progressiveness or regressiveness of the heterogeneity of conditional treatment effect above δ.

Definition 5 *If $\Lambda_{\delta}(y)$ is strictly decreasing (increasing) in y, the treatment effects above δ at lower margin are progressively (regressively) heterogeneous in Y_0. If $\Psi_{\delta}(y)$ is strictly decreasing (increasing) in y, the treatment effects above δ at upper margin are progressively (regressively) heterogeneous in Y_0. If both $\Lambda_{\delta}(y)$ and $\Psi_{\delta}(y)$ are decreasing (increasing) in y, the treatment effects above δ are progressively (regressively) heterogeneous in Y_0.*

Although intuitive, these concepts are not implementable because, as is $F_\Delta(\delta)$, $\Lambda_\delta(y)$ and $\Psi_\delta(y)$ are only partially identified. So, we cannot observe $\Lambda_\delta(y)$ and $\Psi_\delta(y)$ but do observe the bounds of them. Theorem 7 provides their bounds. See Appendix A for the proof.

Theorem 7 *Let Y_1 and Y_0 be continuous random variables. For y such that $F_0(y) \in (0,1)$, $\Lambda_\delta^L(y) \le \Lambda_\delta(y) \le \Lambda_\delta^U(y)$, and $\Psi_\delta^L(y) \le \Psi_\delta(y) \le \Psi_\delta^U(y)$, where*

$$\begin{aligned}
\Lambda_\delta^L(y) &= \min\left\{\max\left\{\frac{\sup_{x\le y}\{F_0(x)-F_1(\delta+x)\}}{F_0(y)},0\right\},1\right\},\\
\Lambda_\delta^U(y) &= \max\left\{\min\left\{\frac{\inf_{x\le y}\{F_0(x)-F_1(\delta+x)\}+1}{F_0(y)},1\right\},0\right\},\\
\Psi_\delta^L(y) &= \min\left\{\max\left\{\frac{\sup_{x\ge y}\{F_0(x)-F_1(\delta+x)\}-F_0(y)}{1-F_0(y)},0\right\},1\right\},\\
\Psi_\delta^U(y) &= \max\left\{\min\left\{1+\frac{\inf_{x\ge y}\{F_0(x)-F_1(\delta+x)\}}{1-F_0(y)},1\right\},0\right\}.
\end{aligned}$$

and these bounds are sharp.

The interpretation of these bounds is not as intuitive as that of Q^L or Q^U. $\Lambda_\delta^L(y)$ is the minimum fraction of subpopulation whose pre-treatment outcomes are no more than y who benefit at least δ from the treatment. Consider $y = F_0^{-1}(p)$. Then, at least $100{*}\Lambda_0^L\left(F_0^{-1}(p)\right)$ % of low p-th quantile people (measured by pre-treatment outcomes) will benefit from the treatment. $\Lambda_0^U\left(F_0^{-1}(p)\right)$ is the maximum fraction of benefiters. Therefore at least $100{*}\left\{1-\Lambda_0^U\left(F_0^{-1}(p)\right)\right\}$ % of low p-th quantile people (in terms of pre-treatment outcomes) will lose from the treatment. The interpretation of $\Psi_\delta^L(y)$ and $\Psi_\delta^U(y)$ are analogous except that these two consider the sub-population whose pre-treatment outcomes are equal to or higher than y.

These bounds can be used to test for the homogeneity. Since the homogeneity of treatment effects implies $\Lambda_\delta(y) = \Psi_\delta(y) \in \{0,1\}$ for all δ and y, if we find $\left[\Lambda_\delta^L(y), \Lambda_\delta^U(y)\right]$ or $\left[\Psi_\delta^L(y), \Psi_\delta^U(y)\right]$ that is a proper subset of $[0,1]$, it is decisive evidence against the homogeneity.

In addition, these bounds can be used to test for the weak homogeneity. Figure 12 shows the concept.

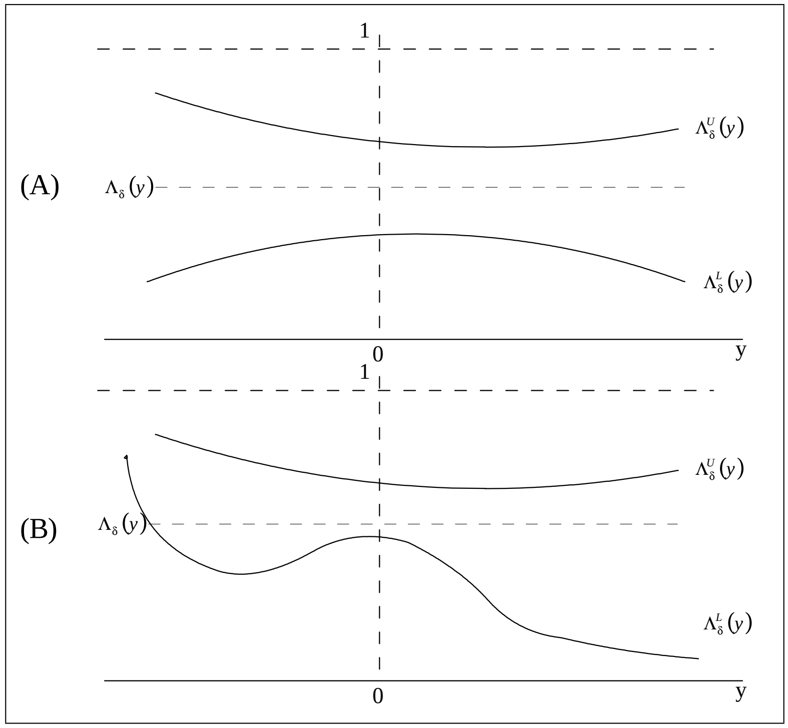


Figure 12. Concept of Testing for 'Weakly Homogeneous Treatment Effects'

Weak homogeneity implies $\Lambda_\delta(y)$ and $\Psi_\delta(y)$ should be flat in y for all δ. Figure 12 (A) provides a type of $\Lambda_\delta^L(y)$ and $\Lambda_\delta^U(y)$ that allow the treatment effects to be weakly homogeneous. The $\Lambda_\delta^L(y)$ and $\Lambda_\delta^U(y)$ shown in Figure 12 (A) can be bounds for a flat $\Lambda_\delta(y)$ such as the one shown in the figure. On the contrary, The $\Lambda_\delta^L(y)$ and $\Lambda_\delta^U(y)$ in Figure 12 (B) cannot be bounds for a flat $\Lambda_\delta(y)$. Therefore, if we see $\Lambda_\delta^L(y)$ and $\Lambda_\delta^U(y)$ such as the ones in Figure 12 (B) then the treatment effects cannot be weakly homogeneous. More formally, if we find some δ at which

$$\sup_y \Lambda_\delta^L(y) > \inf_y \Lambda_\delta^U(y) \text{ or } \sup_y \Psi_\delta^L(y) > \inf_y \Psi_\delta^U(y),$$

then the weak homogeneity assumption is rejected.

What these bounds cannot tell us is whether the conditional heterogeneity of treatment effects above δ is progressive, regressive, or neither. It's because these bounds do not degenerate typically. Nonetheless, I suggest to use these bounds as a crude measure of progressiveness of the conditional treatment effects because the decreasing or increasing behavior of those bounds can be of interest to the policy maker. The reason is, recalling $\Lambda_\delta^U(y)$ is the maximum probability that an individual benefits more than or equal to δ when this person's pre-treatment outcome is no greater than y, we can see if $\Lambda_\delta^U(y)$ is increasing then the policy change can be favorable to the rich. Similar arguments apply to the other three concepts. For this use, I suggest the following definitions.

Definition 6 *The treatment effects above δ are progressively heterogeneous at the extreme if $y < y'$ implies*

$$\text{either } \left[\Lambda_\delta^L(y'), \Lambda_\delta^U(y')\right] \subset \left[\Lambda_\delta^L(y), \Lambda_\delta^U(y)\right] \text{ or } \left[\Psi_\delta^L(y'), \Psi_\delta^U(y')\right] \subset \left[\Psi_\delta^L(y), \Psi_\delta^U(y)\right]$$

but does not imply

$$\Lambda_\delta^L(y') > \Lambda_\delta^L(y),\ \Lambda_\delta^U(y') > \Lambda_\delta^U(y),\ \Psi_\delta^L(y') > \Psi_\delta^L(y),\ \text{or } \Psi_\delta^U(y') > \Psi_\delta^U(y)$$

for all δ. The 'regressiveness at the extreme' can be defined in a similar manner.

The expression 'at the extreme' is to emphasize the idea that the progressiveness or regressiveness happens at the boundaries of the identification regions of $\Lambda_\delta(y)$ and $\Psi_\delta(y)$, not $\Lambda_\delta(y)$ and $\Psi_\delta(y)$ themselves.

The nonparametric estimators I use in this chapter for the bounds for $\Lambda_\delta(y)$ and $\Psi_\delta(y)$ are;

$$\begin{aligned}
\Lambda_{\delta n}^L(y) &= \min\left\{\max\left\{\frac{\sup_{x\le y}\{F_{0n}(x) - F_{1n}(\delta+x)\}}{F_{0n}(y)}, 0\right\}, 1\right\}, \\
\Lambda_{\delta n}^U(y) &= \max\left\{\min\left\{\frac{\inf_{x\le y}\{F_{0n}(x) - F_{1n}(\delta+x)\} + 1}{F_{0n}(y)}, 1\right\}, 0\right\}, \\
\Psi_{\delta n}^L(y) &= \min\left\{\max\left\{\frac{\sup_{x\ge y}\{F_{0n}(x) - F_{1n}(\delta+x)\} - F_{0n}(y)}{1 - F_{0n}(y)}, 0\right\}, 1\right\}, \\
\Psi_{\delta n}^U(y) &= \max\left\{\min\left\{1 + \frac{\inf_{x\ge y}\{F_{0n}(x) - F_{1n}(\delta+x)\}}{1 - F_{0n}(y)}, 1\right\}, 0\right\},
\end{aligned}$$

for y such that $0 < F_{n0}(y) < 1$.

The asymptotic theory of these estimators remains to be established. I will leave it for future research.

Empirical Results

Bounds for Quantiles

Figure 11 shows the bounds for the quantile function of the treatment effects in reading. As described earlier, I considered two sets of treatment and control pairs: One S and R pair and the other A and R pair. Throughout this section, Q_{SR} is the quantile function of treatment effects when the treatment and control pair is S and R and Q_{AR} is the quantile function of the treatment for the pair of A and R. The bounds for the quantile functions were estimated at 101 equally scattered points in $[0.015, 0.985]$ including the two boundary points.[8] I will use the subscript n to denote the estimates. All of the confidence intervals presented are for the true quantile at the 0.95 confidence level. Black color is used for the S and R pair and red (gray on black and white prints) for the A and R pair. Solid lines are the estimates of the Q^L and Q^U functions and dashed lines present the confidence intervals.

For CI_{FP}, I used nonparametric density estimation with the bandwidth 1.34*interquartile range. $b_n = n_1^{-1/3}$ for CI_{FP} and $d_n = n_1^{1/3}$ for CI_{NEW}, where n_1 is the number of data points in Y_1. To avoid unnecessary complexity on the graphs below, I presented CI_{NEW} only. For all cases, CI_{FP} were slightly tighter than CI_{NEW}. However, based on the results on Chapter III on the skewed distributions in that CI_{NEW} worked better than CI_{FP}, and

[8] To simplify wordings, I will use 'for all p' to mean 'for all points at which the bounds are estimated'.

observing the density estimates of the distribution of student scores are skewed unimodal in Ding and Lehrer (2005), I present CI_{NEW} only. I also define $\hat{p}_U = \sup_{p\in[0,1]} Q_n^U(p) \leq 0$ and $\hat{p}_L = \inf_{p\in[0,1]} Q_n^L(p) \geq 0$ indicating $\hat{p}_U$ and $\hat{p}_L$ are the estimators of p_U and p_L. However, no statistical inference on them were made.

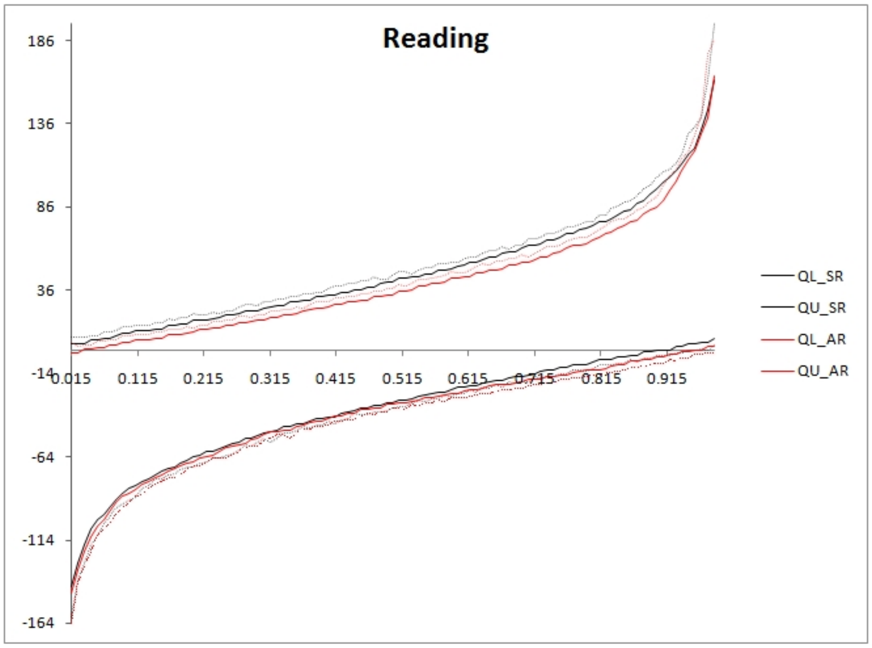


Figure 13. Estimates of the bounds of Q_{TE} in Reading

First I show Q_{SR} and Q_{AR} in Figure 13. $Q_{SRn}^U > Q_{ARn}^U$ for all p and, moreover, Q_{SRn}^U for $p \leq 0.9077$ are above the upper bound of confidence intervals for Q_{AR} implying Q_{SR}^U is significantly larger than Q_{AR}^U pointwise. Although it is not clear on the graph, $Q_{SRn}^L \geq Q_{ARn}^L$ and Q_{ARn}^L are smaller than the lower bounds of CI for Q_{SR} for $p > 0.55$, which implies Q_{SR}^L are significantly larger than Q_{AR}^L pointwise from $p \geq 0.694$. $\hat{p}_L$ and $\hat{p}_U$ are shown in Table 23.

Table 23. $\hat{p}_L$ and $\hat{p}_U$ for Reading

Reading	$\hat{p}_{L,SR}$	$\hat{p}_{L,AR}$	$\hat{p}_{U,SR}$	$\hat{p}_{U,AR}$
All sample	0.8977	0.9559	<0.015	0.0538

About 10% of the population will gain for sure if the policy change is $R \to S$ while

$R \to A$ benefits only 4.5% of the population. In terms of losses, 5.4% of the population are estimated to lose from $R \to A$ whereas less than 1.5% lose from $R \to S$. Based on the discussion of Section V.2.1, S appears to be preferred by policy makers if S and A are equally costly.

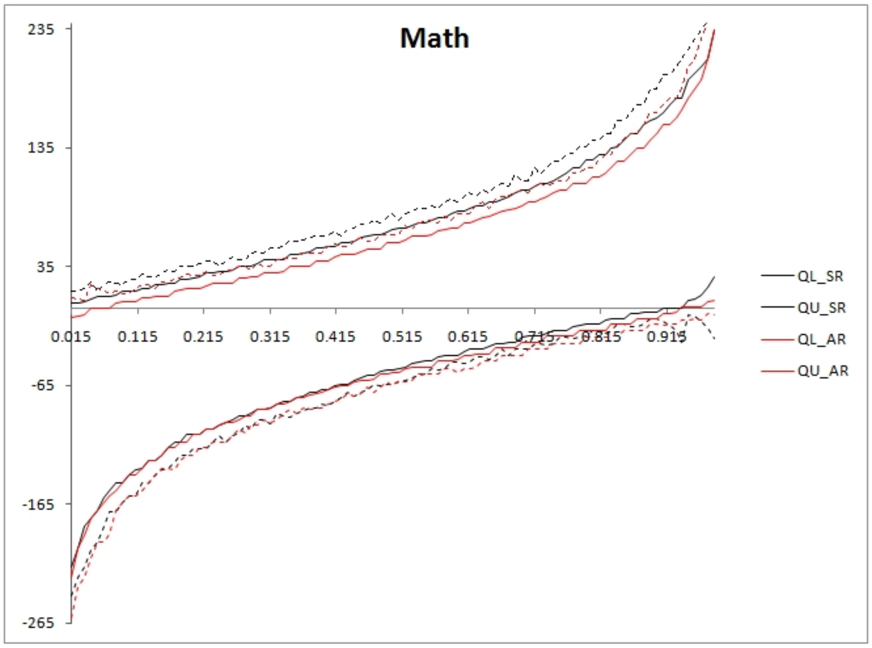


Figure 14. Estimates of the bounds of Q_{TE} in Math

For math, we have Figure 14. Similar patterns appear but, in a sense, with less certainty. Although $Q^U_{SRn} > Q^U_{ARn}$, Q^U_{SRn} at no p exceed the upper bound of the confidence intervals of Q_{AR} constantly. $\hat{p}_L$ and $\hat{p}_U$ are in Table 24. Although it is less convincing than the case of reading, focusing on Q^U_n, we still can say S might be preferred.

Table 24. $\hat{p}_L$ and $\hat{p}_U$ for Math

Math	$\hat{p}_{L,SR}$	$\hat{p}_{L,AR}$	$\hat{p}_{U,SR}$	$\hat{p}_{U,AR}$
All sample	0.9074	0.9365	<0.015	0.0829

To see how the heterogeneity in treatment effects differs with students' characteristics, I split the whole sample into eight subgroups according to student's sex, race, and socioeconomic status. The socioeconomic status is measured by whether or not student received free lunch. The subgroups[9] and related descriptive statistics are as follows.

Table 25. Subgroup Categories

Group		Class	Reading # of obs.	Reading Average	Math # of obs.	Math Average
First Catego-rization	(Male, Free Lunch: MP)	S	418	428.8	421	475.1
		A	494	425.4	501	468.0
		R	495	422.4	502	467.1
	(Male, No Free Lunch: MR)	S	471	446.8	478	502.8
		A	541	441.1	556	490.5
		R	520	437.6	524	486.1
	(Female, Free Lunch: FP)	S	400	435.2	406	482.5
		A	513	429.0	518	473.6
		R	452	429.0	458	476.2
	(Female, No Free Lunch: FR)	S	443	449.9	450	501.1
		A	468	446.8	473	499.9
		R	511	449.3	520	502.4
Second Catego-rization	(White, Free Lunch: WP)	S	399	432.5	402	481.2
		A	444	427.7	446	476.1
		R	426	428.3	428	475.2
	(White, No Free Lunch: WR)	S	784	449.4	798	503.9
		A	889	444.6	903	496.6
		R	908	444.5	917	494.7
	(Non White, Free Lunch: NP)	S	419	431.4	425	476.3
		A	563	426.8	573	466.8
		R	521	423.3	532	468.4
	(Non White, No Free Lunch: NR)	S	130	441.3	130	490.5
		A	120	437.1	126	482.1
		R	123	435.4	127	490.5

Throughout the rest of chapter, I will abbreviate notations for each group. Except for NR, all subgroups have enough observations for Small, Regular-With-Aide, and Regular classes. Due to data limitations, we need be careful when interpreting the result of NR. For NR, I report the confidence intervals only for $p \in [0.1, 0.9]$. Table 26 provides the summary for $\hat{p}_L$ and $\hat{p}_U$. It should be emphasized that these numbers are all estimates without testing if one is significantly different from another.

By the same argument for the entire sample case, S may be desired by policy makers for all subsample categories in reading. However, the actual $\hat{p}_L$ and $\hat{p}_U$ are quite different across subgroups implying that different subgroups may benefit by different amount. In the first categorization, if the policy change is $R \rightarrow S$, then almost 13% of MR will have non-

[9] As is evident, the subgroups are not mutually exclusive. I did not construct mutually exclusive subgroups due to the concern about the number of observations in each subgroup. Another method that can be used is a nonparametric or semiparametric estimation with covariates.

Table 26. $\hat{p}_L$ and $\hat{p}_U$ for Subgroups

	Reading				Math			
	$\hat{p}_{L,SR}$	$\hat{p}_{L,AR}$	$\hat{p}_{U,SR}$	$\hat{p}_{U,AR}$	$\hat{p}_{L,SR}$	$\hat{p}_{L,AR}$	$\hat{p}_{U,SR}$	$\hat{p}_{U,AR}$
MP	0.8589	0.8977	<0.015	0.0344	0.888	0.8783	0.0247	0.1023
MR	0.8298	0.9462	<0.015	0.0344	0.8201	0.9074	<0.015	0.0441
FP	0.9074	0.9462	<0.015	0.1217	0.8783	0.9462	0.0441	0.112
FR	0.9462	0.985	0.0441	0.0926	0.9462	0.9462	0.0732	0.0829
WP	0.9074	0.9559	0.0247	0.0829	0.8977	0.9074	0.0538	0.1023
WR	0.888	0.9656	0.0247	0.0635	0.8686	0.9365	0.0247	0.0538
NP	0.8589	0.9171	<0.015	0.0441	0.888	0.9074	0.0344	0.112
NR	0.8492	0.8977	<0.015	0.1411	0.8492	0.9365	0.0829	0.112
* All	0.8977	0.9559	<0.015	0.0538	0.9074	0.9365	<0.015	0.0829

negative benefits for sure while the proportion is only about 5% in FR. When the change is $R \rightarrow A$ then now MP has the largest fraction who benefit. About 10% of MP will benefit for certain while FR is still the least. Moreover, at least 4.4% of FR will lose under the change of $R \rightarrow S$. This measures the heterogeneous effects of CSR. If instead the change is $R \rightarrow A$, then FP has the highest fraction of people who will get hurt while more than 12% of FP will lose for sure.

In the second categorization, the $S \rightarrow R$ seems most beneficial to NR and NP. About 15% of each group will benefit for certain. In terms of the fractions who lose for certain, still NP and NR are the best. Less than 1.5% of each group lose from the change. However, the disadvantage of NR is the change from $S \rightarrow A$ in terms of the fraction of people who lose for certain. The $S \rightarrow R$ is most beneficial to MR in the first categorization in terms of both the fractions benefiting and losing. The effect on NR is controversial. NR have the biggest benefiting fraction yet the biggest losing fraction of people suffer from for certain loss as well.

Since $\sup_p Q^L(p) > \inf_p Q^U(p)$ provides evidence of non-homogeneous treatment effects, Table 27 and 28 show whether the homogeneous treatment hypothesis is admissible or not. All cases considered revealed $\sup_p Q_n^L(p) > \inf_p Q_n^U(p)$, which suggests the effects of both S and A treatments are not likely to be homogeneous.

Table 27. $\sup_p Q_n^L(p)$ and $\inf_p Q_n^U(p)$ (Reading)

	Reading			
	$\max Q_{SRn}^L$	$\min Q_{SRn}^U$	$\max Q_{ARn}^L$	$\min Q_{ARn}^U$
MP	7	4	2	-2
MR	12	4	3	-6
FP	17	2	3	-6
FR	4	-4	0	-8
WP	5	0	2	-11
WR	7	0	2	-4
NP	10	4	12	-2
NR	10	1	15	-8
* All	7	4	2	-2

Table 28. $\sup_p Q_n^L(p)$ and $\inf_p Q_n^U(p)$ (Math)

	Math			
	$\max Q_{SRn}^L$	$\min Q_{SRn}^U$	$\max Q_{ARn}^L$	$\min Q_{ARn}^U$
MP	11	0	13	-14
MR	43	5	26	0
FP	12	0	5	-15
FR	8	-6	5	-17
WP	10	-12	13	-12
WR	29	0	7	-8
NP	26	0	6	-17
NR	11	-43	5	-55
* All	26	4	6	-8

The statistical tests for the homogeneous treatment effect hypothesis is the test of

$$H_0 \;:\; F^L(\delta) = 0 \text{ and } F^U(\delta) = 1 \text{ for some } \delta,$$

$$H_A \;:\; F^L(\delta) > 0 \text{ or } F^U(\delta) < 1 \text{ for all } \delta,$$

where $F^L(\cdot)$ and $F^U(\cdot)$ are the bounds for $F_\Delta(\delta)$ defined in Chapter II. We can test the pointwise version of these hypotheses by extending Chapter II to develop a test for the homogeneous treatment effect hypothesis or by using the confidence intervals established in the chapter directly. Let

$$H_0' \;:\; F^L(\delta) = 0 \text{ and } F^U(\delta) = 1 \text{ for a given } \delta,$$

$$H_A' \;:\; F^L(\delta) > 0 \text{ or } F^U(\delta) < 1 \text{ for a given } \delta.$$

To test H_0', we can simply construct two one-sided confidence intervals for $F^L(\delta)$

Table 29. Tests for H_0'

	Reading				Math			
	Reject?(SR)	δ	Reject?(AR)	δ	Reject?(SR)	δ	Reject?(AR)	δ
MP	?	.	?	.	?	.	Yes	1.28
MR	Yes	8.4	Yes	3.2	Yes	11.87	Yes	0.98
FP	?	.	Yes	-2.8	Yes	95	Yes	-5.98
FR	Yes	-0.98	Yes	-4	Yes	-3.92	?	.
WP	Yes	6	?	.	Yes	1.42	Yes	-2.16
WR	Yes	2.68	?	.	Yes	12.34	?	.
NP	?	.	?	.	?	.	Yes	0.7
NR	?	.	Yes	0.5	Yes	-8.27	Yes	-7.09
All	?	.	Yes	1.32	Yes	4.06	Yes	0

and $F^U(\delta)$ and see if they contain 0 and 1 respectively. Table 29 presents the results of this tests. "?" in the table means none of δ's considered rejected H_0'. Under the change of $R \to S$, MR, FR, WP, and WR rejected the hypothesis of homogeneity of treatment effects in reading. In math under $R \to S$, MR, FP, FR, WP, WR, and NR rejected it. When $R \to A$ is considered, MR, FP, FR, and NR rejected the hypothesis in reading. In math under $R \to S$, almost all of subgroups rejected it except for FR and WR.

Estimation of Bounds for the Conditional Distribution of Treatment Effects upon Pre-Treatment Outcomes

Next, I plotted the estimated bounds for $\Lambda_{SR,\delta}(y)$, $\Lambda_{AR,\delta}(y)$, $\Psi_{SR,\delta}(y)$, $\Psi_{AR,\delta}(y)$ against quantiles of Y_0 for $\delta = 0$. Since $\delta = 0$, the conditional probability considered here is the probability that an individual will benefit from each treatment when the individual is at the lower margin ($\Lambda_\delta(y)$) or upper margin ($\Psi_\delta(y)$). By construction, $\Lambda^L_{n,\delta}(y)$ and $\Lambda^U_{n,\delta}(y)$ (solid lines) are more informative at lower quantiles and $\Psi^L_{n,\delta}(y)$ and $\Psi^U_{n,\delta}(y)$ (dashed lines) are more informative at upper quantiles. In the estimation, $\Lambda^U_{nSR,0}(y) = \Lambda^U_{nAR,0}(y) = 1$ and $\Psi^L_{nSR,0}(y) = \Psi^L_{nAR,0}(y) = 0$ for almost all y considered. Therefore, $\Lambda^L_{n,0}(y)$ (solid lines at the lower part of each graph) and $\Psi^L_{n,0}(y)$ (solid lines at the upper part of each graph) deserve more attention. As in the previous section, the black color indicates S treatment and A treatment uses the red color. From here on, I will omit the subscript 0 from above functions for the sake of notational convenience as long as the omission does not disrupt understanding.

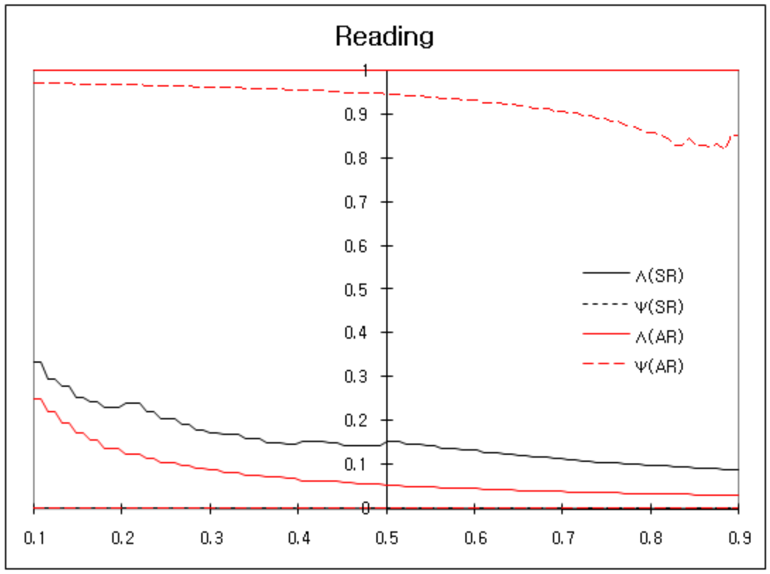


Figure 15. Estimates of bounds for $\Lambda_0(y)$ and $\Psi_0(y)$

First, let us look at the reading (Figure 15). Λ^L_{nSR} and Λ^L_{nAR} are above from 0 and Ψ^U_{nAR} are substantially smaller than 1. From this, the rejection of the homogeneity assumption for both of SR and AR seems likely. $\Lambda^L_{nSR}(y) > \Lambda^L_{nAR}(y)$ for all y considered, implying that certain fraction of students will benefit from $R \to S$ are more than from $R \to A$. Both of the treatments S and A appear to be progressively heterogeneous at the extreme. $\Lambda^L_{nSR}(y)$ at $y = F_0^{-1}(0.1)$ is 0.335, meaning that if a student's score in a regular sized class is below the 10 percentile of pre-treatment outcome, the probability that his score will increase after the CSR is 0.335. In other way of interpreting it is about 33.5% of students in the 10th percentile or lower will be better after the CSR.

The bounds are roughly constant or downward sloping. So, the treatment effects are progressively heterogeneous at the extreme, which implies more of less-able students will benefit non-negative amounts for sure from both S and A than more-able students. Also the maximum possible proportion of benefiters is larger between less-able students than between more-able students. This is along the lines of the findings of Ding and Lehrer (2005).

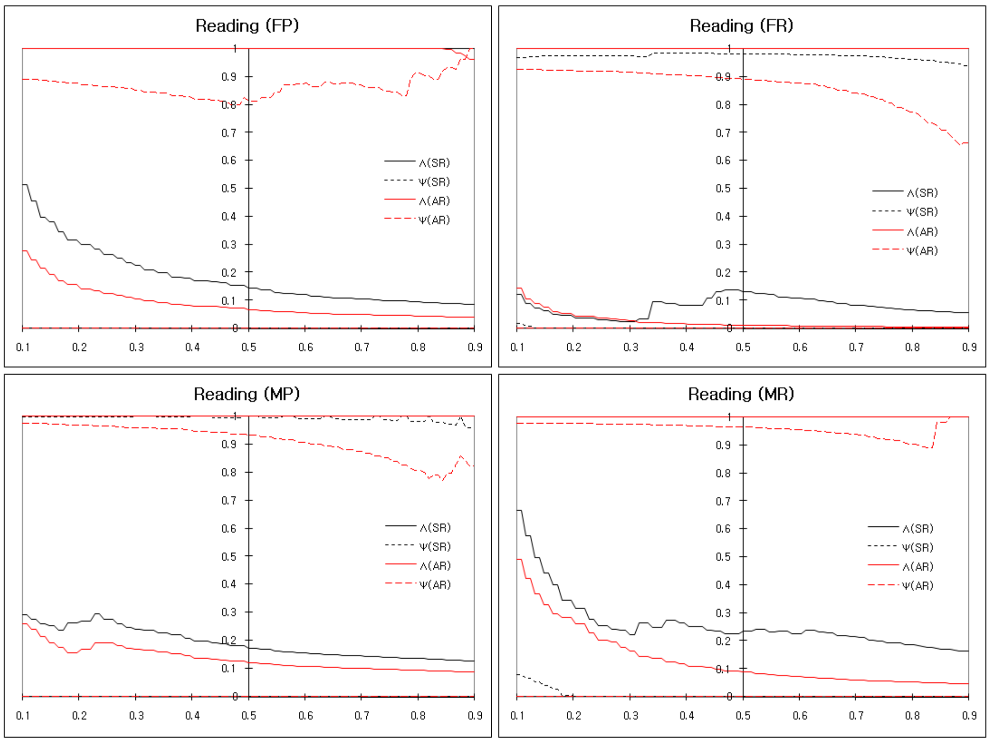


Figure 16. Bounds for $\Lambda(y)$ and $\Psi(y)$ (Category 1)

The bounds for Λ and Ψ for subgroups are presented in Figure 16. The homogeneity seems to be rejected in both of the treatments in all subgroups, as well. $\Lambda^L_{nSR}(y) \geq \Lambda^L_{nAR}(y)$ for all y in all subgroups except for the low quantile of FR group, which implies more students benefit for sure from $R \rightarrow S$ then from $R \rightarrow A$. The actual $\Lambda^L_{nSR}(y)$ differ greatly across subgroups from the highest of 0.62 for $F_0^{-1}(0.1)$ of MR to the lowest of 0.024 for $F_0^{-1}(0.3)$ of FR. The downward sloping $\Lambda^L_n(y)$ in FP and MR means less-able students have more for certain fractions of benefiters in those subgroups. In FR, however, the fraction of for certain benefiters decreases initially as y increases and increases from

$y = F_0^{-1}(0.3)$ to $y = F_0^{-1}(0.3)$ suggesting Ding and Lehrer (2005)'s finding may not hold for all subgroups.

Since $1 - \Psi_{nAR}^{U}$ is the minimum fraction of people who will lose from $R \rightarrow A$, the upward sloping Ψ_{nAR}^{U} in FP means more-able students in FP will lose less from $R \rightarrow A$. This upward tendency of Ψ_{n}^{U} at the upper quantiles exists in all subgroups. A is less desirable than S in this regard, too.

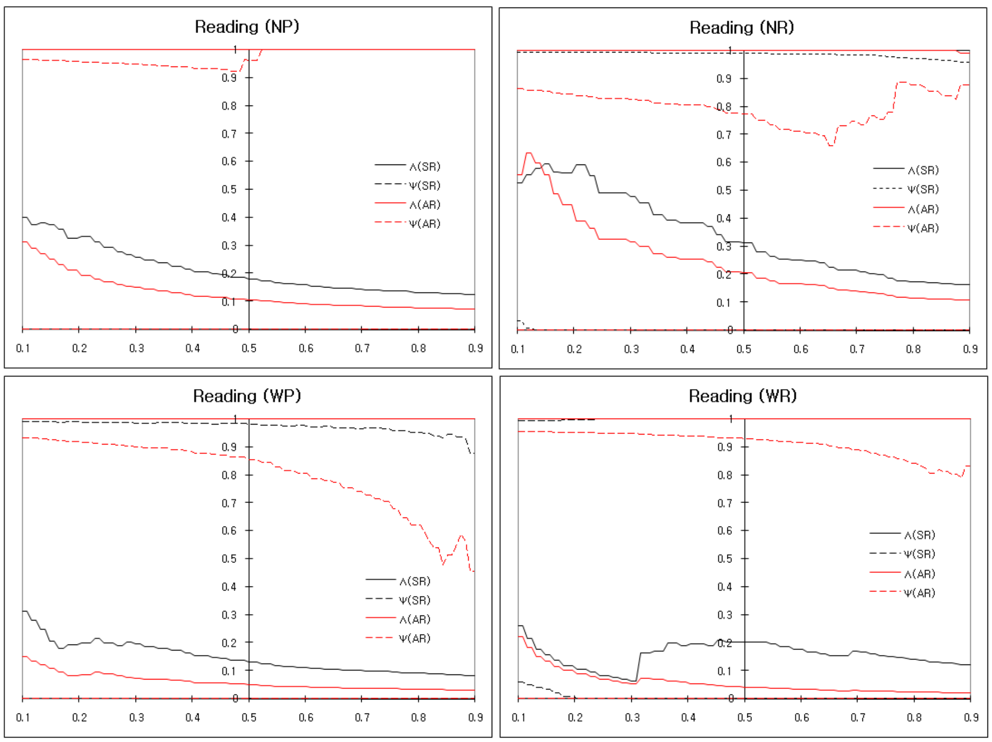


Figure 17. Bounds for $\Lambda(y)$ and $\Psi(y)$ (Category 2)

Figure 17 is the second categorization. Non-trivial fractions of more-able students in WP and NR appear to lose from $R \rightarrow S$ as can be seen from the pattern of $\Psi_{nSR}^{U}(y)$. $\Psi_{nSR}^{U}\left(F_0^{-1}(0.9)\right)$ is 0.957 in NP and 0.877 in WP. Therefore, at least 4.3% of highly able students in NP and 12.3% of those in WP will experience losses from the CSR. The $\Lambda_{nSR}^{L}(y)$ for WR provides more evidence that Ding and Lehrer (2005)'s finding may not hold in some subgroups.

NR having the largest $\Lambda_n^L(y)$ overall suggests that they are gain the largest benefits. The C group in general has larger $\Lambda_n^L(y)$ than the W group. This supports the finding that non-white students benefited more from the CSR. None of the two treatment effects look either progressive or regressive at the extreme. We can analyze the treatment effects in math similarly. See Appendix C for the graphs and discussions in math.

We can assess the heterogeneity in the treatment effects in a different manner. I presented graphs of $\Lambda_\delta(y)$ and $\Psi_\delta(y)$ at fixed y's but at different δ's. To understand the use of them, let us consider two students whose scores are $F_0^{-1}(0.2)$ and $F_0^{-1}(0.8)$ and the functions $\Lambda_\delta\left(F_0^{-1}(0.2)\right)$ and $\Psi_\delta\left(F_0^{-1}(0.8)\right)$. By definition, $\Lambda_\delta\left(F_0^{-1}(0.2)\right)$ is the probability that a student whose pre-treatment score is no more than $F_0^{-1}(0.2)$ (a low ability student) gains δ when the class environment changes and $\Psi_\delta\left(F_0^{-1}(0.8)\right)$ is the probability that a student whose pre-treatment score is no less than $F_0^{-1}(0.8)$ (a high ability student) gains δ when the class environment changes.

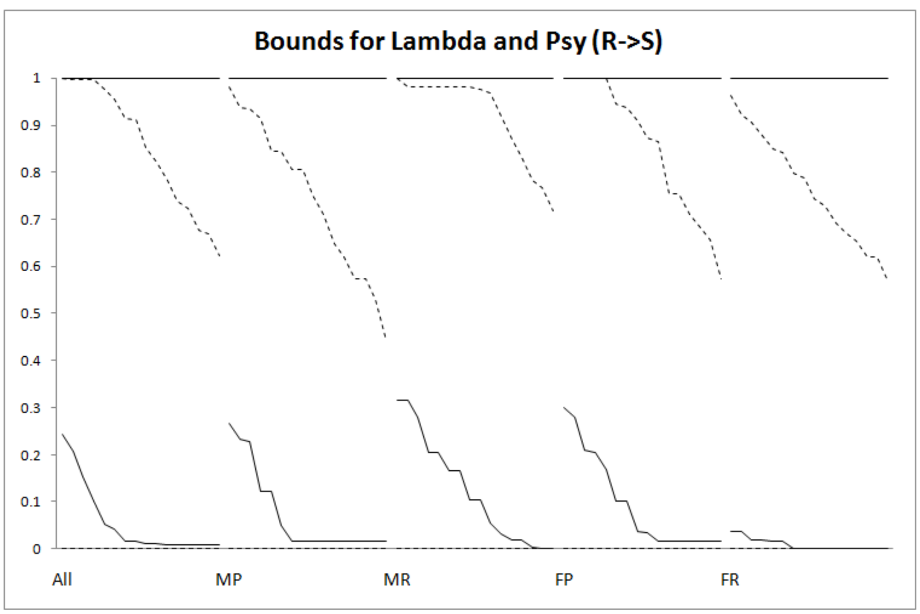


Figure 18. Bounds for $\Lambda(y)$ and $\Psi(y')$ (Category 1)

Figure 18 presents the estimated bounds for $\Lambda_\delta\left(F_0^{-1}(0.2)\right)$ and $\Psi_\delta\left(F_0^{-1}(0.8)\right)$ in reading for $\delta \in [0, 15]$ in $R \to S$ for the whole sample (denoted by 'All') and the subgroups in Category 1. $F_0^{-1}(0.2)$ in reading was 410 and $F_0^{-1}(0.8)$ was 456. The solid lines are the bounds for $\Lambda_\delta\left(F_0^{-1}(0.2)\right)$ and the dashed lines are for $\Psi_\delta\left(F_0^{-1}(0.8)\right)$. For the entire sample, the lower bounds for $\Psi_\delta\left(F_0^{-1}(0.8)\right)$ are 0 for all δ and the upper bounds are downward sloping from 0 to 0.621. Since the lower bounds are 0, the $F_0^{-1}(0.8)$ student may not benefit at all. $\Psi_{\delta n}^U\left(F_0^{-1}(0.8)\right) = 0.621$ for $\delta = 15$ means that only 62.1% of high ability students (students whose scores are higher than or equal to 456) will gain an additional 15 points or more when the class size reduces. For the $F_0^{-1}(0.2)$ student, the lower bounds at $\delta = 1$ is 0.207, which means that at least 20.7% of low ability students (those whose scores are lower than or equal to 410) will score at least 1 point more after the CSR. When we consider $\delta = 10$, the lower bound for $\Lambda_\delta\left(F_0^{-1}(0.2)\right)$ is 0.094, which means at least 9.4% of low ability students will score at least 10 points more after CSR.

When we look into subgroups, the gains vary. The upper bounds for $\Psi_\delta\left(F_0^{-1}(0.8)\right)$ for MP at $\delta = 0$ is 0.981. 98.1% of high ability students in MP group gain non negative amounts from the treatment and the remaining 1.9% will score less after the CSR. The upper bounds for $\Psi_\delta\left(F_0^{-1}(0.8)\right)$ for FR at $\delta = 0$ is 0.963. 4.7% of high ability FR students will score less after CSR. FR is the subgroup that benefits the least of the four, also, in terms of the minimum certainty gainers. The lower bounds for $\Lambda_\delta\left(F_0^{-1}(0.2)\right)$ at $\delta = 0$ is only 0.038, meaning that just 3.8% of low ability students in FR will gain non-negative amounts for sure from CSR. The MR, on the other hand, shows the highest lower bounds for $\Lambda_\delta\left(F_0^{-1}(0.2)\right)$ at $\delta = 0$, which is 0.316.

Figure 19 shows the results for the other categorization. The upper bound for $\Psi_\delta\left(F_0^{-1}(0.8)\right)$ for WP at $\delta = 0$ is 0.951. At least about 4.5% of high ability students in the WP will hurt from the CSR. $\Psi_\delta\left(F_0^{-1}(0.8)\right)$ being 0.972 for NR at $\delta = 0$ means at least

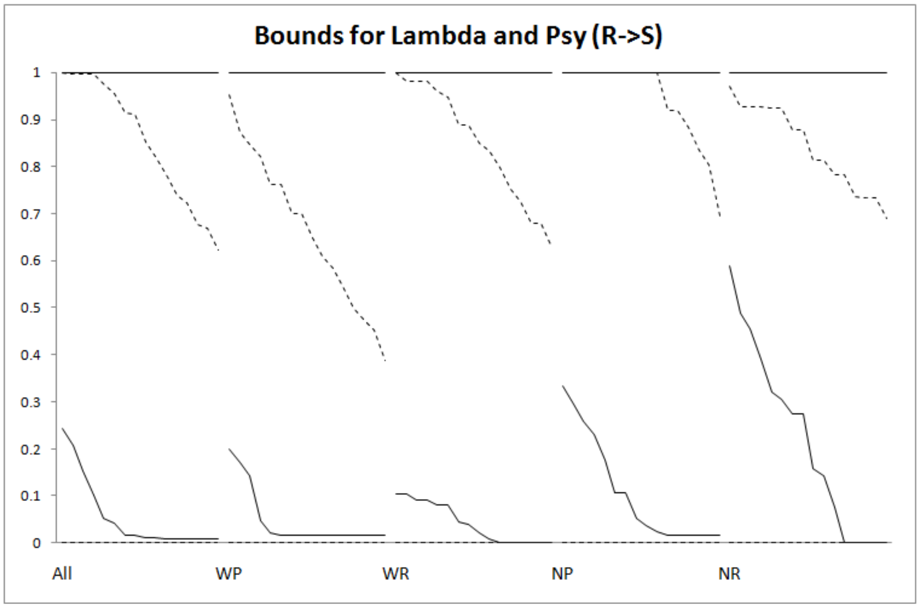


Figure 19. Bounds for $\Lambda(y)$ and $\Psi(y')$ (Category 2)

2.8% of high ability students in the NR group will lose. 16% of NP and 17% of the WP will gain additional 15 points or more after the CSR.

Conclusion

In this chapter, I investigated the potential heterogeneous treatment effects of class size reduction (CSR) experimented in Project STAR conducted by the Tennessee State Department of Education. This chapter strengthened the findings of Ding and Lehrer (2005) on the heterogeneous treatment effects. However, differently from them, I found the direction of heterogeneity is not as simple as they concluded based on quantile regression results.

To make my findings most general and robust, I used the bounding approach and the new inference method in Chapter II. This approach does not assume any type of dependence structure between Y_1 and Y_0 whereas the quantile regressions or QTE implicitly assume the perfect positive dependence. I furthered the ideas in Chapter III to develop tools to investigate heterogeneity more closely by defining weak homogeneity and conditional heterogeneity of treatment effects upon pre-treated outcomes and by providing the bounds of conditional distribution of treatment effects on the pre-treatment outcomes.

Applying those theories, I constructed the bounds for unconditional distribution and the heterogeneity of conditional treatment effects on pre-treatment outcomes and found the following: i) there is strong evidence of heterogeneity of treatment effects in both A and S treatments for all subgroups; ii) the features of heterogeneity differ across subgroups; iii) the distributional impact of treatment effects should not be ignored; iv) the conditional heterogeneity on pre-treatment outcomes are meaningful for policy evaluation and implementation.

To conclude, the partial identification approach (bounding approach in this case), when applied to Project STAR, indicates that the heterogeneity of treatment effect is important and should be addressed in empirical work. Further, the results suggest that we may not restrict our research on a particular dependence structure without any theoretical basis in general by using ATE or QTE.

There are many things to be done. In the econometric theoretical point of view, the construction of the uniformly valid confidence interval for the true $F_{\Delta}(\delta)$, the asymptotics of the estimators for the bounds of $\Lambda_{\delta}(y)$ and $\Psi_{\delta}(y)$, and the inference on the conditional distributions of treatment effects on observables are in progress. Extension of the present chapter are also needed in regards to the empirical application of Project STAR. My current research dealt with just kindergarten students' performances. To extend this chapter to higher year's accomplishment, we need to know how to handle the selection on unobservables. Fan and Wu (2007) may be useful on this issue. In addition, if we may investigate potential results of policy implementation prior to the implementation of it by conducting a randomized experiment. Manski (1997, 2003) called this a 'mixing problem'. The inference on the mixing problem is also of interest. All of these topics remain for the future work.

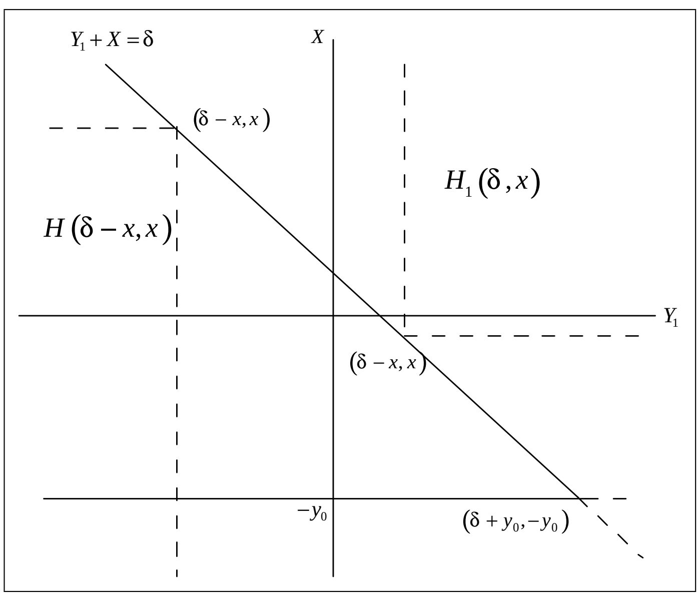
Appendix A. Technical Proof

I will provide the bounds for $\Pr\left[\Delta \leq \delta, Y_0 \leq y_0\right]$ then the bounds for $\Lambda_\delta\left(y_0\right)$ can be computed by

$$\Lambda_\delta\left(y_0\right) = \Pr\left[\Delta > \delta | Y_0 \leq y_0\right] = 1 - \frac{\Pr\left[\Delta \leq \delta, Y_0 \leq y_0\right]}{\Pr\left[Y_0 \leq y_0\right]}.$$

Let $X = -Y_0$. Then $\Pr\left[\Delta \leq \delta, Y_0 \leq y_0\right] = \Pr\left[\Delta \leq \delta, X \geq -y_0\right]$. Define $W\left(u, v\right) = \max\left\{u + v - 1, 0\right\}$ and $M\left(u, v\right) = \min\left\{u, v\right\}$. Also define $H\left(y_1, x\right) = \Pr\left[Y_1 \leq y_1, X \leq x\right]$.

$\Pr\left[\Delta \leq \delta, X \geq -y_0\right]$ is the H-volume of upper left half plane surrounded by the lines $Y_1 + X = \delta$ and $X = -y_0$.



As in Nelson (1993), $\Pr\left[\Delta \leq \delta, X \geq -y_0\right]$ is bounded from below by $\sup_{x \geq -y_0} H\left(\delta - x, x\right)$.

Therefore,

$$
\begin{aligned}
& \Pr\left[\Delta \leq \delta, X \geq -y_0\right] \\
\geq\ & \max\left\{\sup_{x \geq -y_0} H\left(\delta - x, x\right) - H\left(\delta - x, -y_0\right), 0\right\} \\
\geq\ & \max\left\{\sup_{x \geq -y_0} \left[\max\left\{F_1\left(\delta - x\right) + F_X\left(x\right) - 1, 0\right\} - \min\left\{F_1\left(\delta - x\right), F_X\left(-y_0\right)\right\}\right], 0\right\} \\
=\ & \max\left\{\sup_{y \leq y_0} \left[\max\left\{F_1\left(\delta + y\right) - F_0\left(y\right), 0\right\} - \min\left\{F_1\left(\delta + y\right), 1 - F_0\left(y_0\right)\right\}\right], 0\right\} \\
=\ & \max\left\{\sup_{y \leq y_0} \left\{F_1\left(\delta + y\right) - F_0\left(y\right) - 1 + F_0\left(y_0\right)\right\}, 0\right\}
\end{aligned}
$$

For the upper bound, we know

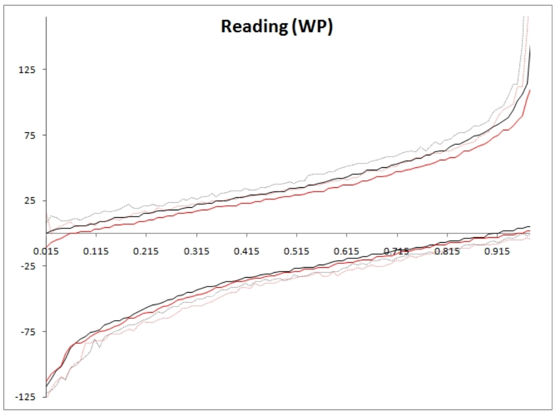
$$
\begin{aligned}
& \Pr\left[\Delta \leq \delta, X \geq -y_0\right] \\
\leq\ & \inf_{x \geq -y_0} \left\{1 - F_X\left(-y_0\right) - H_1\left(\delta, x\right)\right\} \\
=\ & \inf_{x \geq -y_0} \left\{1 - F_X\left(-y_0\right) - \left\{1 - F_1\left(\delta - x\right) - F_X\left(x\right) + H\left(\delta - x, x\right)\right\}\right\} \\
\leq\ & \inf_{x \geq -y_0} \left\{-F_X\left(-y_0\right) + F_1\left(\delta - x\right) + F_X\left(x\right) - \max\left\{F_1\left(\delta - x\right) + F_X\left(x\right) - 1, 0\right\}\right\} \\
=\ & \inf_{y \leq y_0} \left\{F_0\left(y_0\right) + F_1\left(\delta + y\right) - F_0\left(y\right) - \max\left\{F_1\left(\delta + y\right) - F_0\left(y\right), 0\right\}\right\} \\
=\ & F_0\left(y_0\right) + \min\left\{\inf_{y \leq y_0} \left\{F_1\left(\delta + y\right) - F_0\left(y\right)\right\}, 0\right\}
\end{aligned}
$$

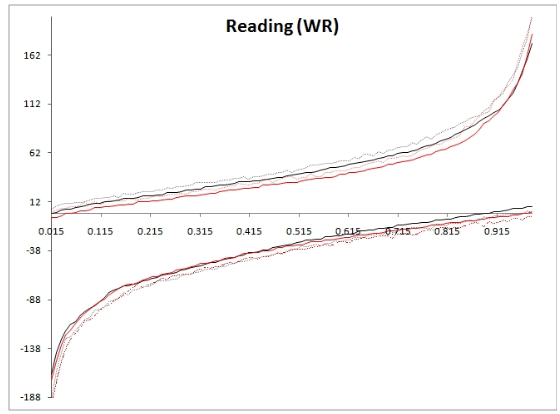
The proof for $\Psi_\delta\left(y_0\right)$ require the bounds for $\Pr\left[\Delta \leq \delta, Y_0 \geq y_0\right]$. The proof is analogous except that we have to start with

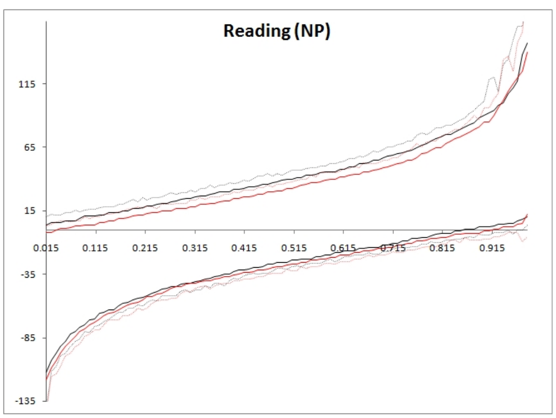
$$
\begin{aligned}
& \sup_{x \leq -y_0} H\left(\delta - x, x\right) \\
\leq\ & \Pr\left[\Delta \leq \delta, X \geq -y_0\right] \\
\leq\ & \inf_{x \leq -y_0} \left\{F_X\left(x\right) + H\left(\delta - x, -y_0\right) - H\left(\delta - x, -x\right)\right\}.
\end{aligned}
$$

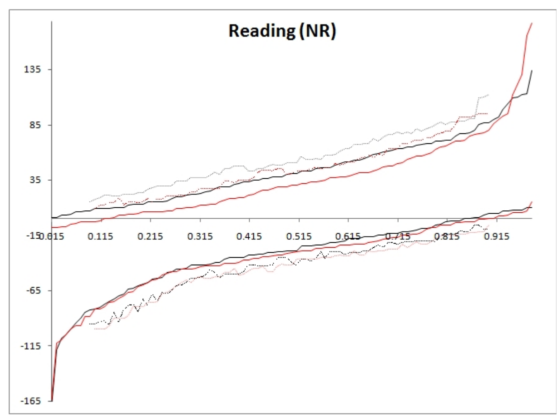
The reason that these bounds are sharp is analogous to Lemma 1.

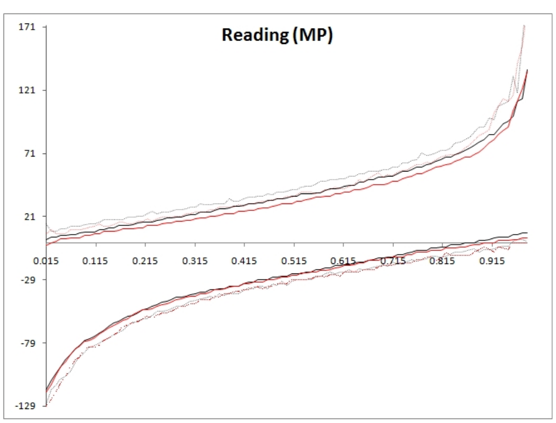
Appendix B. Graphs of the Estimates for the Bounds for $Q_{\mathbf{TE}}\,(p)$ for Subgroups

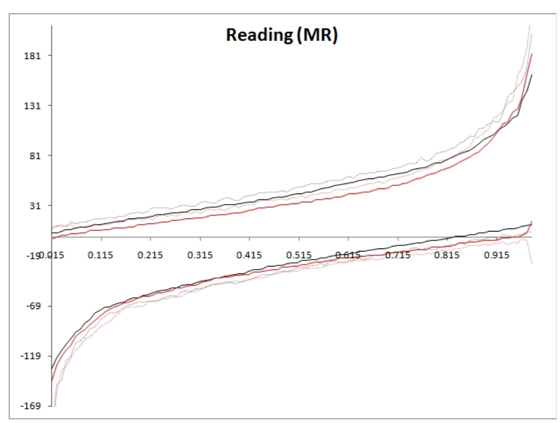


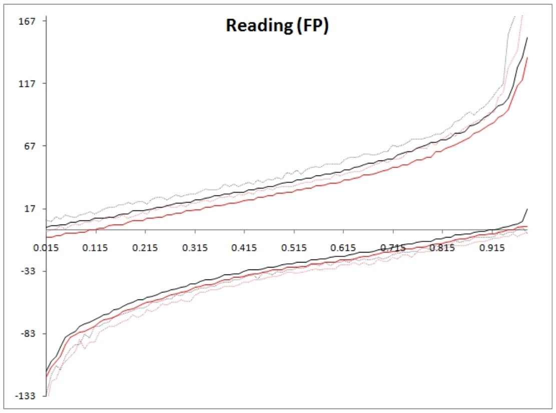




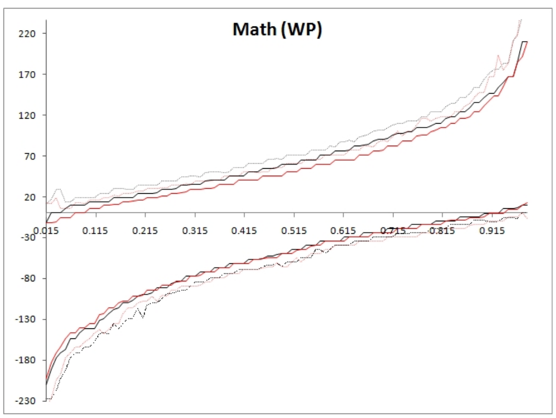


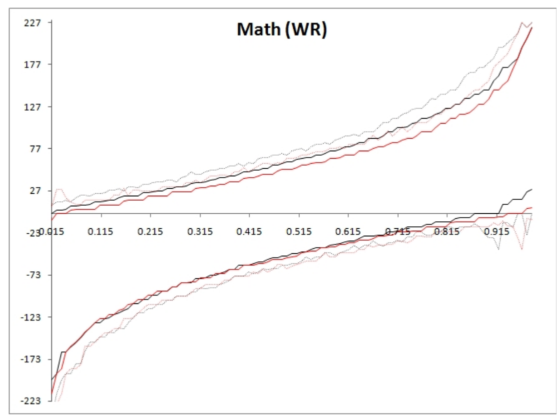


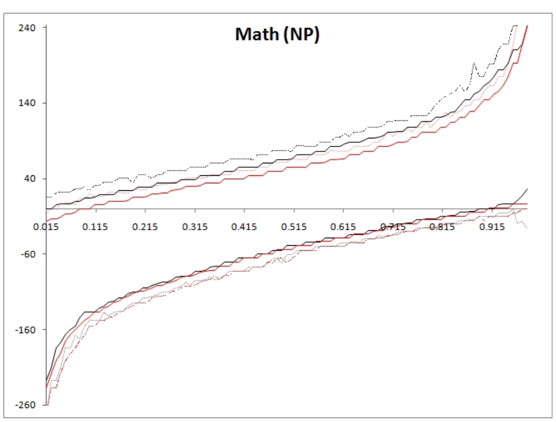


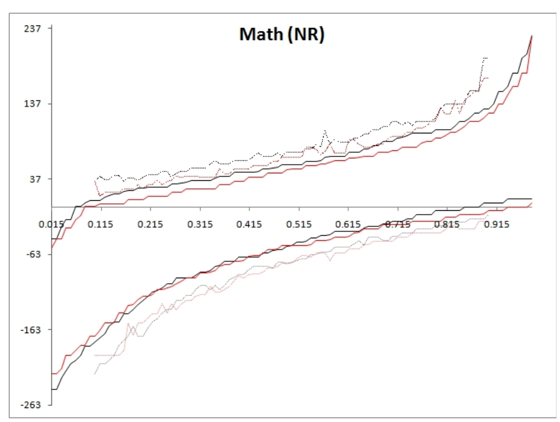

Reading (FP)
167
117
67
17
-33
-83
-133
0.015
0.115
0.215
0.315
0.415
0.515
0.615
0.715
0.815
0.915

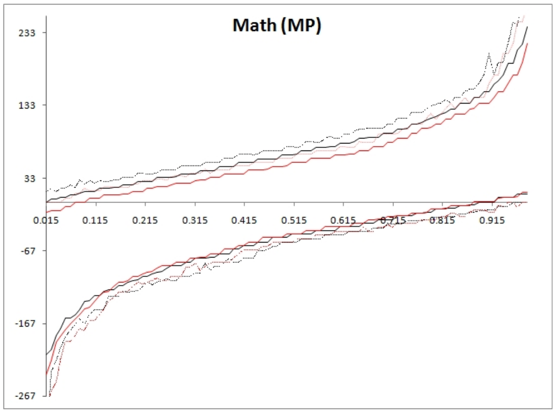

Reading (FR)
163
113
63
13
-37
-87
-137
-187
0.015
0.115
0.215
0.315
0.415
0.515
0.615
0.715
0.815
0.915

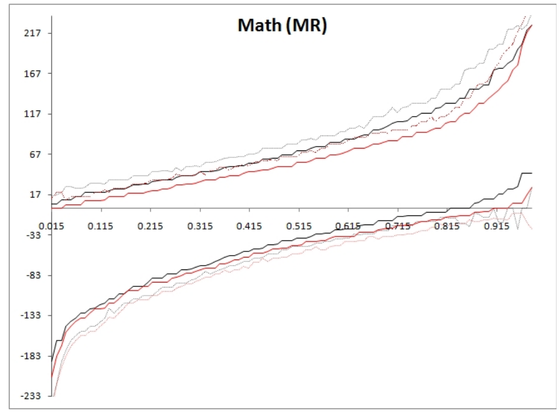

Math (WP)
220
170
120
70
20
-30
-80
-130
-180
-230
0.015
0.115
0.215
0.315
0.415
0.515
0.615
0.715
0.815
0.915

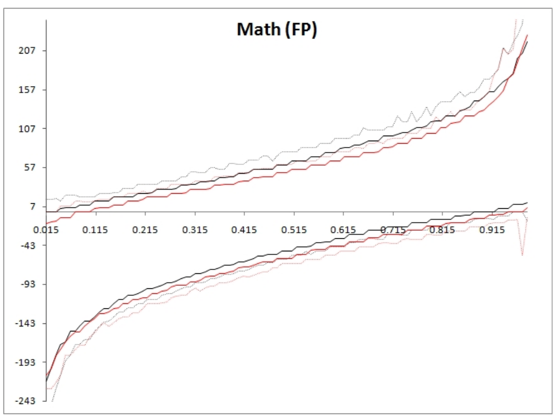

Math (WR)
227
177
127
77
27
-23
-73
-123
-173
-223
0.015
0.115
0.215
0.315
0.415
0.515
0.615
0.715
0.815
0.915

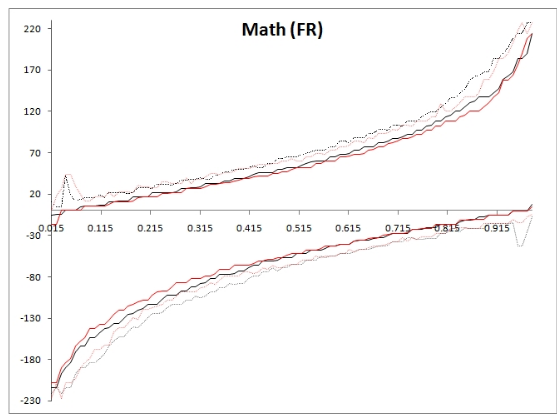

Math (NP)
240
140
40
-60
-160
-260
0.015
0.115
0.215
0.315
0.415
0.515
0.615
0.715
0.815
0.915


Math (NR)
237
137
37
-63
-163
-263
0.015
0.115
0.215
0.315
0.415
0.515
0.615
0.715
0.815
0.915

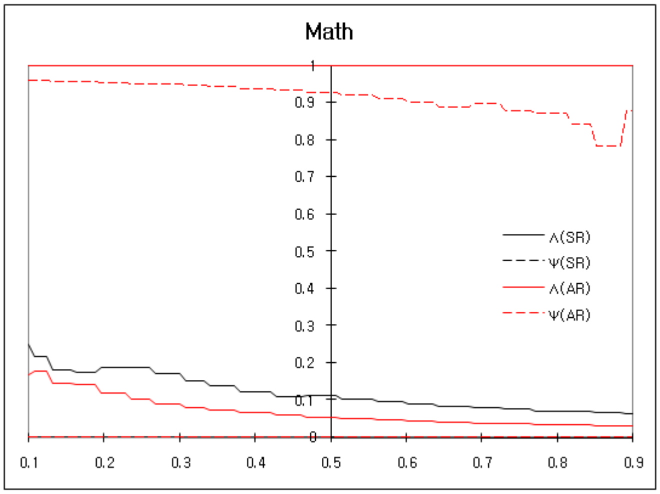

Math (MP)
233
133
33
-67
-167
-267
0.015
0.115
0.215
0.315
0.415
0.515
0.615
0.715
0.815
0.915

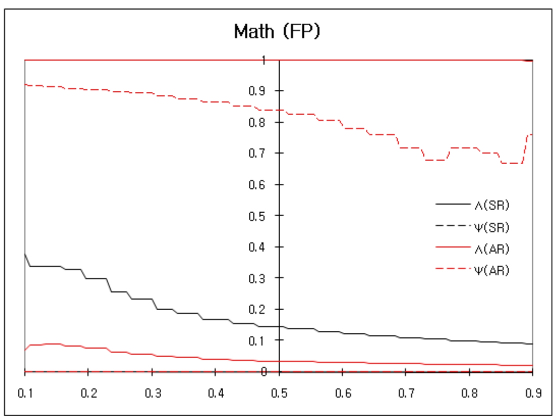

Math (MR)
217
167
117
67
17
-33
-83
-133
-183
-233
0.015
0.115
0.215
0.315
0.415
0.515
0.615
0.715
0.815
0.915

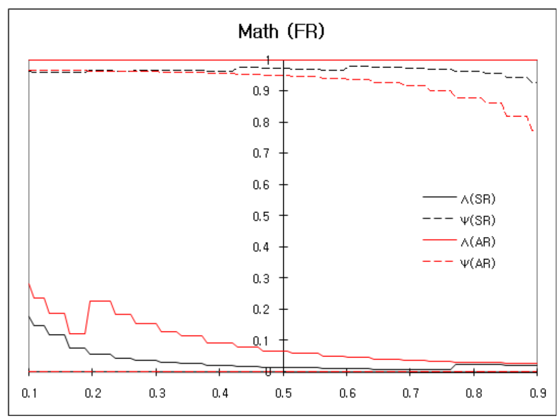

Math (FP)
207
157
107
57
7
-43
-93
-143
-193
-243
0.015
0.115
0.215
0.315
0.415
0.515
0.615
0.715
0.815
0.915

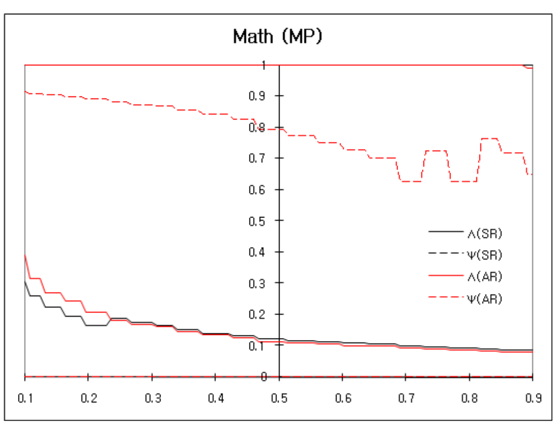

Math (FR)
220
170
120
70
20
-30
-80
-130
-180
-230
0.015
0.115
0.215
0.315
0.415
0.515
0.615
0.715
0.815
0.915

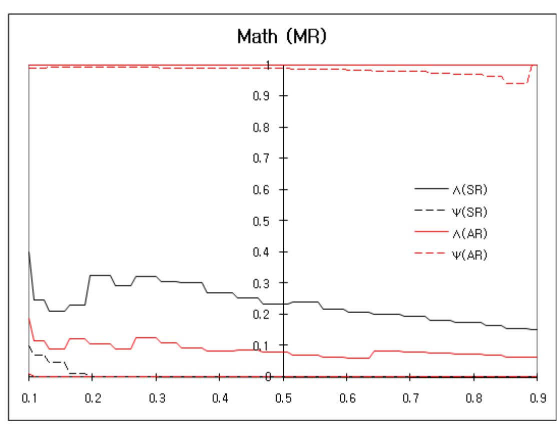
Appendix C. Graphs of $\Lambda_0(y)$ and $\Psi_0(y)$ in Math

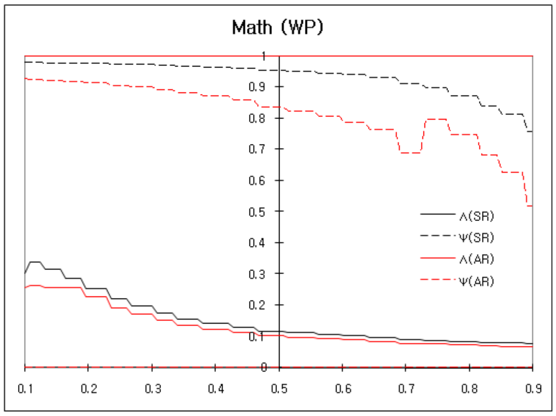


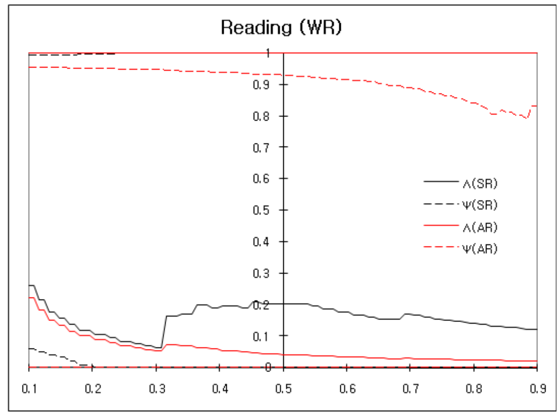


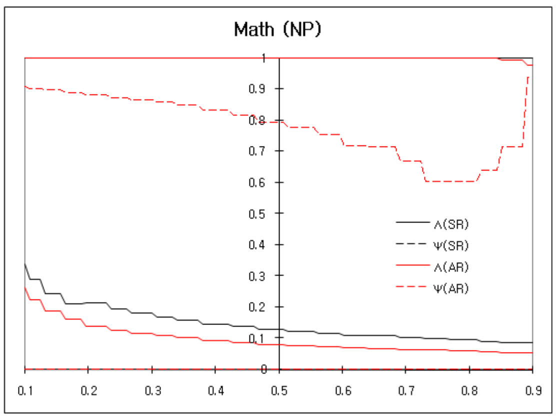


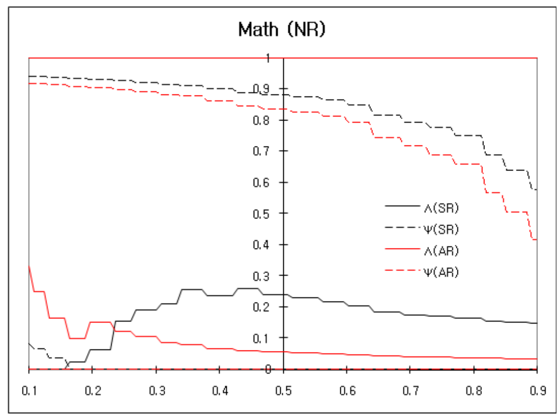





Math (WP)
Λ(SR)
Ψ(SR)
Λ(AR)
Ψ(AR)


Reading (WR)
Λ(SR)
Ψ(SR)
Λ(AR)
Ψ(AR)


Math (NP)
Λ(SR)
Ψ(SR)
Λ(AR)
Ψ(AR)


Math (NR)
Λ(SR)
Ψ(SR)
Λ(AR)
Ψ(AR)

CHAPTER VI

Conclusion

It has been widely agreed that individuals receive different benefits from treatments such as social programs or policy implementation (i.e. treatment effects are not homogeneous over individuals). While observed heterogeneity in individual characteristics accounts for part of the heterogeneity in treatment effects, heterogeneous treatment effects remain even after controlling for observed heterogeneity in individual characteristics. Consequently, the average treatment effect (ATE), or the ATE conditional upon some observed covariates, provides too little information when we need to know the distributional aspect of treatment effects.

A direct look into the distribution of treatment effects is a solution. However, of the two parts of information we need in order to find the distribution of treatment effects - information on the outcomes with and without the treatment -, only one is observed for a given individual. This fundamental missing data problem precludes the point-identification of distribution of treatment effects even in a randomized experiment without imposing additional (often non-reputable) assumptions such as rank preservation assumption.

Instead of imposing such assumptions, this dissertation takes the bounding approach. As is thoroughly investigated in Williamson and Downs (1990), the distribution of a random variable which is an arithmetic function of two random variables is bounded by two probability distributions. Because the treatment effect is defined by the difference between two outcomes, the distribution of treatment effects is also bounded, thus partially identified.

The first essay develops the econometric tools for inference on partially identified parameters. The second essay examines the partial identification of the distribution of treatment effects extensively. There, we develop the asymptotic distribution of the estimator for each bound. We leave the statistical inference on the distribution of treatment effects for future work and, instead, focus on the quantile function of treatment effects in the third essay. Since this function is the inverse of the distribution of treatment effects, it is partially identified as well. We develop the asymptotics of the estimators for the bounds of quantiles of treatment effects and apply the tools developed in the first essay for the statistical inference. We also propose a new technique of statistical inference here. The last essay applies the theories and techniques developed in other three essays to Project STAR and finds substantial evidence of heterogeneous treatment effects of class size reduction (CSR) on kindergarten students' achievement.

The dissertation contributes to related streams of literature in several ways. The confidence sets in the first essay are applicable not only to the identification of treatment effects but also to various economic problems in which parameters are only partially identified. The partial identification of the distribution or quantile function of treatment effects provides robust techniques that enable a deeper understanding of heterogeneous treatment effects. The concept of distribution of treatment effects conditional on pre-treatment outcome in the fourth essay enables us to answer the question of "who gains from treatment, who loses, and by how much" more explicitly. The empirical study on Project STAR provides more convincing evidence of the heterogeneity of the effects of CSR. In addition, the fourth essay shows that the pattern of heterogeneity due to unobservable factors, such as individual ability, may differ in relation with observable characteristics such as gender or ethnicity. This suggests that earlier research on Project STAR, which found that the CSR benefited low socioeconomic group more in reading and higher ability students overall, may not fully capture the complex patterns of heterogeneity in treatment effects.

Many things remain undone. The asymptotics in the second essay necessitates the advancement of the first essay to allow for cases where asymptotic distribution of model parameters is discontinuous. The inference on the distribution of treatment effects conditional upon covariates is a generalization that is also being considered. It is of interest because the distribution of treatment effects conditional upon covariates can allow us to better understand the pattern of heterogeneity of treatment effects. The very idea presented in the third essay will be applied to what Manski called the 'mixing problem'. Additionally, the assumptions that underpin the inference techniques developed in the second essay will be relaxed in future work. This relaxation will enable the inference on the true distribution of treatment effects conditional on pre-treatment outcomes. Finally, the application of these theories to existing social program data is another interest of the author.

BIBLIOGRAPHY

[1] Aakvik, A. , J. Heckman, and E. Vytlacil (2003). "Treatment Effects for Discrete Outcomes When Responses to Treatment Vary Among Observationally Identical Persons: An Application to Norwegian Vocational Rehabilitation Programs." Forthcoming in *Journal of Econometrics*.

[2] Abadie, A., J. Angrist, and G. Imbens (2002). "Instrumental Variables Estimation of Quantile Treatment Effects." *Econometrica* 70, 91-117.

[3] Alsina, C. (1981). "Some Functional Equations in the Space of Uniform Distribution Functions." *Equationes Mathematicae* 22, 153-164.

[4] Andrews, D. W. K. (2000). "Inconsistency of the Bootstrap When a Parameter is on the Boundary of the Parameter Space." *Econometrica* 68, 399-405.

[5] Andrews, D. W. K. and P. Guggenberger (2005a), "The Limit of Exact Size and a Problem with Subsampling and with the m Out of n Bootstrap," Unpublished Manuscript, Cowles Foundation, Yale University.

[6] Andrews, D. W. K. and P. Guggenberger (2005b), "Hybrid and Size-corrected Subsampling Methods," Unpublished Manuscript, Cowles Foundation, Yale University.

[7] Andrews, D. W. K. and P. Guggenberger (2005c), "Applications of Subsampling, Hybrid, and Size-correction Methods," Cowles Foundation Discussion Paper No. 1608, Yale University.

[8] Andrews, D. W. K. and P. Guggenberger (2007), "Validity of Subsampling and "Plug-in Asymptotic" Inference for Parameters Defined by Moment Inequalities," Unpublished Manuscript, Cowles Foundation, Yale University.

[9] Beresteanu, A. and F. Molinari (2006), "Asymptotic Properties for a Class of Partially Identified Models," Manuscript.

[10] Beran, R. (1997). "Diagnosing Bootstrap Success." *Ann. Inst. Statist. Math.* 49, 1-24.

[11] Berry, S.T. (1992), "Estimation of a Model of Entry in the Airline Industry", *Econometrica* 60, 889-917.

[12] Bickel, P. J., F. Götze and W. R. Zwet (1997). "Resampling Fewer than n Observations: Gains, Losses, and Remedies for Losses." *Statistica Sinica* 7, 1-31.

[13] Bickel, P. J., F. and A. Sakov (2005). "On the Choice of m in the m out of n Bootstrap and its Application to Confidence Bounds for Extreme Percentiles." Working paper.

[14] Biddle, J., L. Boden and R. Reville (2003). "A Method for Estimating the Full Distribution of a Treatment Effect, With Application to the Impact of Workfare Injury on Subsequent Earnings." Mimeo.

[15] Bitler, M., J. Gelbach, and H. W. Hoynes (2006). "What Mean Impact Miss: Distributional Effects of Welfare Reform Experiments." Forthcoming in *American Economic Review.*

[16] Black, D. A. , J. A. Smith, M. C. Berger and B. J. Noel (2003). "Is the Threat of Reemployment Services More Effective Than the Services Themselves? Experimental Evidence From the UI System." *American Economic Review* 93(3), 1313-1327.

[17] Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2006). "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds." Forthcoming in *Econometrica*.

[18] Bourke, S. (1998) "How Smaller Is Better: Some Relationships Between Class Size, Teaching Practices and Student Achievement." *American Educational Research Journal*, 23: 558—571, 1986.

[19] Bresnahan, T. and P. Reiss (1991), "Empirical models of discrete games," *Journal of Econometrics* 48, 57–82.

[20] Bugni, F. A. (2007), "Bootstrap Inference in Partially Identified Models," mimeo, Northwestern University.

[21] Cambanis, S., G. Simons and W. Stout (1976). "Inequalities for $\mathcal{E}k(X, Y)$ when the Marginals are Fixed." *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 36, 285-294.

[22] Carneiro, P. , K. T. Hansen, and J. Heckman (2003). "Estimating Distributions of Treatment Effects With an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice." *International Economic Review* 44(2), 361-422.

[23] Chernozhukov, V. and C. Hansen (2005). "An IV Model of Quantile Treatment Effects." *Econometrica* 73, 245-261.

[24] ____________________________ (2006). "Instrumental Quantile Regression Inference for Structural and Treatment Effect Models." *Journal of Econometrics* 132. 491-525.

[25] Chernozhukov, V., Lee, S., and Rosen, A. M. (2007). "Inference on Intersection Bounds." Unpublished working paper.

[26] Chernozhukov, V., H. Hong, and E. Tamer (2004), "Inference on Parameter Sets in Econometric Models," Working Paper, MIT.

[27] Chernozhukov, V., H. Hong, and E. Tamer (2007). "Parameter Set Inference in a Class of Econometric Models," *Econometrica* 75, 1243-1284.

[28] Chernozhukov, V., Lee, S., and Rosen, A. M. (2007). "Inference on Intersection Bounds." Unpublished working paper.

[29] Ciliberto, F. and E. Tamer (2004), "Market Structure and Multiple Equilibria in Airline Markets," mimeo, Northwestern University.

[30] Csrögő, M. (1983). *Quantile Processes with Statistical Applications*, Philadelphia, PA.

[31] Davidson, R. and J. G. Mackinnon (2004). *Econometric Theory and Method*. Oxford University Press.

[32] Dehejia, R. (1997). "A Decision-theoretic Approach to Program Evaluation." Ph.D. Dissertation, Department of Economics, Harvard University.

[33] Dehejia, R. and S. Wahba (1999). "Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs." *Journal of the American Statistical Association* 94, 1053-1062.

[34] Dennis, B. D. (1986), "Effects of Small Class Size (1:15) on the Teaching/Learning Process in Grade Two." Dissertation. Tennessee State University, 177.

[35] Denuit, M. , C. Genest, and E. Marceau (1999). "Stochastic Bounds on Sums of Dependent Risks." *Insurance: Mathematics and Economics* 25, 85-104.

[36] Ding, W. and S. Lehrer (2004). "Estimating Dynamic Treatment Effects from Project STAR." Mimeo.

[37] Ding, W. and S. Lehrer (2005). "Class Size and Student Achievement: Experimental Estimates of Who Benefits and Who Loses from Reductions." Queen's Economic Department Working Paper No. 1046, Queen's University.

[38] Djebbari, H. and J. A. Smith (2008), "Heterogeneous Impacts in PROGRESA." Institute for the Study of Labor(IZA) Discussion Paper No. 3362.

[39] Doksum, K. (1974). "Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case." *Annals of Statistics* 2, 267-277.

[40] Embrechts, P., A. Hoeing, and A. Juri (2003). "Using Copulae to Bound the Value-at-Risk for Functions of Dependent Risks." *Finance & Stochastics* 7(2), 145-167.

[41] Evertson, C., and C. Randolph (1998), "Teaching Practices and Class Size: A New Look at an Old Issue." *Peabody Journal of Education* 67, 85-105.

[42] Frank, M. J. , R. B. Nelsen, and B. Schweizer (1987). "Best-Possible Bounds on the Distribution of a Sum—a Problem of Kolmogorov." *Probability Theory and Related Fields* 74, 199-211.

[43] Fan, Y. (2006). "Statistical Inference on the Frechet-Hoeffding Distribution Bounds." Mimeo.

[44] Fan, Y. and J. Wu (2007), “Sharp Bounds on the Distribution of the Treatment effects in Switching Regimes Models,” Unpublished ManusNRipt, Vanderbilt University.

[45] Filby, N., L. Cahen, G. McCutsheon, and D. Kyle. (1980) “What Happens in Smaller Classes? A Summary Report of a Field Study.” San Francisco, CA: Far West Laboratory for Research and Development, 1-21.

[46] Firpo, S. (2004). "Efficient Semiparametric Estimation of Quantile Treatment Effects." Discussion Paper No.04-01, Department of Economics, University of British Columbia.

[47] Firpo, S. (2005). “Efficient Semiparametric Estimation of Quantile Treatment Effects.” Forthcoming in *Econometrica*.

[48] Frölich, Markus and Blaise Melly (2008). “Unconditional Quantile Treatment Effects under Endogeneity.” Institute for the Study of Labor(IZA) Discussion Paper No. 3288.

[49] Glass, G. V., and M. L. Smith. (1978) *Meta-Analysis of Research on the Relationship of Class Size and Achievement*. San Francisco: Far West Laboratory of Educational Research and Development.

[50] __________________________ (1979) "Meta-analysis on Class Size and Achievement." *Ecuational Evaluation and Policy Analysis* 1, 2-16.

[51] Glass, G. V., Leonard Cahan, Mary Lee Smith, and Nikola Filby, School Class Size: Research and Policy Sage Publications (Beverly Hills, London, New Delhi) 1982

[52] Hahn, J. (1998). “On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects.” *Econometrica* 66, 315-331.

[53] Hallinan, M. T. and A. B. Sorensen. (1985) “Ability Grouping and Student Friendships.” *American Educational Research Journal* 22, 485-499.

[54] Hanushek, E. A. (1998), "The Evidence on Class Size." Occasional Paper Number 98-1, W. Allen Wallias Instityte of Political Economy, University of Rochester.

[55] Heckman, J., H. Ichimura, J. Smith, and P. Todd (1998). “Characterizing Selection Bias Using Experimental Data.” *Econometrica* 66, 1017-1098.

[56] Heckman, J. and R. Robb (1985). “Alternative Methods for Evaluating the Impact of Interventions,” in J. Heckman and B. Singer, eds., *Longitudinal Analysis of Labor Market Data*. New York: Cambridge University Press.

[57] Heckman, J. and J. Smith (1993). “Assessing The Case For Randomized Evaluation of Social Programs,” in *Measuring Labour Market Measures: Evaluating the Effects of Active Labour Market Policies*, ed. by K. Jensen and P. K. Madsen. Copenhagen: Danish Ministry of Labor, 35-96.

[58] Heckman, J., J. Smith, and N. Clements (1997). “Making The Most Out Of Programme Evaluations and Social Experiments: Accounting For Heterogeneity in Programme Impacts.” *Review of Economic Studies* 64, 487-535.

[59] Hirano, K., G. W. Imbens, and G. Ridder (2000). “Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score.” *NBER Technical Working Papers* 0251, National Bureau of Economic Research, Inc.

[60] Honore, Bo E. and A. Lleras-Muney (2006). “Bounds in Competing Risks Models and the War on Cancer.” *Econometrica* 74, 1675-1698.

[61] Horowitz, J. L. and C. F. Manski (2000). “Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data.” *Journal of the American Statistical Association* 95, 77-84.

[62] Imbens, G. W. and C. F. Manski (2004). “Confidence Intervals For Partially Identified Parameters.” *Econometrica* 72, 1845–1857.

[63] Imbens, G. W. and W. Newey (2005). “Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity.” Working Paper.

[64] Imbens, G. W. and D. B. Rubin (1997). “*Estimating Outcome Distributions for Compliers in Instrumental Variables Models.*” Review of Economic Studies 64, 555-574.

[65] Jex, S. M. (2002). *Organizational Psychology: A Scientist–Practitioner Approach*. New York: John Wiley & Sons.

[66] Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall/CRC, London.

[67] Kruger, A. B. (1992), "Economic Considerations and Class Size." Working Paper #447, Princeton University.

[68] Krueger A. B. and D. M. Whitmore (2001), “The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence From Project STAR.” *The Economic Journal* 111, 1-28.

[69] Lalonde, R. (1995). “The Promise of Public-Sector Sponsored Training Programs.” *Journal of Economic Perspectives* 9, 149-168.

[70] Lechner, M. (1999). “Earnings and Employment Effects of Continuous Off-the-job Training in East Germany After Unification.” *Journal of Business and Economic Statistics* 17, 74-90.

[71] Lee, L. F. (2002). “Correlation Bounds for Sample Selection Models with Mixed Continuous, Discrete and Count Data Variables.” Manuscript, The Ohio State University.

[72] Lee, M. J. (2005). *MiNRo-Econometrics for Policy, Program, and Treatment Effects*. Oxford University Press.

[73] Lehmann, E. L. (1974). *Nonparametrics: Statistical Methods Based on Ranks.* Holden-Day Inc., San Francisco, California.

[74] Ma, Lingjie, and Roger Koenker (2006). "Quantile Regression Methods for Recursive Structural Equation Models." *Journal of Econometrics* 134. 471–506.

[75] Makarov, G. D. (1981). "Estimates for the Distribution Function of a Sum of two Random Variables When the Marginal Distributions are Fixed." *Theory of Probability and its Applications* 26, 803-806.

[76] Manski, C. F. (1990). "Non-parametric Bounds on Treatment Effects." *American Economic Review, Papers and Proceedings* 80, 319-323.

[77] ___________ (1994). "The Selection Problem," in *Advances in Econometrics, Sixth World Congress* Vol 1, Editor C. Sims, Cambridge University Press.

[78] ___________ (1997a), "The Mixing Problem in Program Evaluation," *Review of Economic Studies* 64, 537-553.

[79] ___________ (1997b). "Monotone Treatment Effect." *Econometrica* 65, 1311-1334.

[80] ___________ (2003). *Partial Identification of Probability Distributions.* Springer-Verlag, New York.

[81] Manski, C. F. and J. Pepper (2000). "Monotone Instrumental Variables: With Application to the Returns to Schooling." *Econometrica* 68, 997-1010.

[82] Manski, C. F. and E. Tamer (2002), "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica* 70, 519-546.

[83] Marron, J. S. and M. P. Wand (1992). "Exact Mean Integrated Squared Error." *The Annals of Statistics* 20, 712-736.

[84] McNeil, A., R. Frey, and P. Embrechts (2005). *Quantitative Risk Management: Concepts, Techniques, and Tools.* Princeton Series in Finance, Springer Boston.

[85] Millimet, D. L. and A. Kumas (2007). "Reassessing the Effects of Bilateral Tax Treaties on US FDI Activity." Unpublish manuscript. Southern Methodist University.

[86] Moon, R. and F. Schorfheide (2007). "A Bayesian Look at Partially-Identified Models," Manuscript, U.Penn.

[87] Nelsen, R. B. (1999). *An Introduction to Copulas.* Springer, New York.

[88] Newey, W. K. (1991), "Uniform Convergence in Probability and Stochastic Equicontinuity." *Econometrica* 59, 1161-1167.

[89] Pakes, A., J. Porter, K. Ho, and J. Ishii (2006), “Moment Inequalities and Their Application,” Working Paper.

[90] Park, S. (2007a), “Confidence Intervals for the Distribution of the Treatment effects,” in progress.

[91] Park, S. (2007b), “The Mixing Problem in Program Evaluation: Inference,” in progress.

[92] Pate-Bain, H., J. Boyd-Zaharias, Van A. Cain, Elizabeth Word, and M. Edward Binkley (1997). *The Student/Teacher Achievement Ratio (STAR) Project, STAR Follow-up Studies 1996-1997*. Health and Education Research Operative Services (HEROS), Inc.

[93] Politis, D. N. and J. P. Romano (1994). “Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions.” *Annals of Statistics* 22, 2031-2050.

[94] Politis, D. N., J. P. Romano, and M. Wolf (1999). *Subsampling*. Springer-Verlag, New York.

[95] Robinson, G. E. (1990), "Synthesis of Research on Class Size." *Educational Leadership* 47, 80-90.

[96] Robinson, G. E. and J. H. Wittebols. (1986) *Class Size Research: A Related Cluster Analysis for Decision Making*. Education Research Service. Inc.: Research Brief, Arlinton, Virgina.

[97] Romano, J. P. and A. M. Shaikh (2005a), “Inference for Identifiable Parameters in Partially Identified Econometric Models,” Unpublished Working Paper, Department of Economics, University of Chicago.

[98] Romano, J. P. and A. M. Shaikh (2005b), “Inference for the Identified Set in Partially Identified Econometric Models,” Unpublished Working Paper, Department of Economics, University of Chicago.

[99] Romano, J. P. and A. M. Shaikh (2006). “Inference for Identifiable Parameters in Partially Identified Econometric Models.” Working Paper.

[100] Rosen, A.M. (2005), “Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities,” mimeo, University College London.

[101] Rosenbaum, P. R. and D. B. Rubin (1983a). “Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome.” *Journal of the Royal Statistical Society, Series B* 45, 212-218.

[102] Rosenbaum, P. R. and D. B. Rubin (1983b). “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika* 70, 41-55.

[103] Rüschendorf, L. (1982). “Random Variables With Maximum Sums.” *Advances in Applied Probability* 14, 623-632.

[104] Schweizer, B. and A. Sklar (1983). *Probabilistic Metric Spaces*. North-Holland, New York.

[105] Shaikh, A. M. and E. Vytlacil (2005). "Threshold Crossing Models and Bounds on Treatment Effects: A Nonparametric Analysis." Working Paper.

[106] Sklar A. (1959). "Fonctions de réartition à n dimensions et leures marges," *Publications de l'Institut de Statistique de L'Université de Paris* 8, 229-231.

[107] Soares, G. (2006), "Inference for Partially Identified Models with Inequality Moment Constraints," Working Paper, Yale.

[108] Stoye, J. (2005). "Partial Identification of Spread Parameters." Working paper.

[109] Stoye, J. (2007), "More on Confidence Intervals for Partially Identified Parameters." Working paper.
[110] Schweizer, B. and A. Sklar (1983). *Probabilistic Metric Spaces*. North-Holland, New York.

[111] Shaikh, A. M. and E. Vytlacil (2005). "Threshold Crossing Models and Bounds on Treatment Effects: A Nonparametric Analysis." Working Paper.
[112] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London.

[113] Tamer, E. (2003), "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," Review of Economic Studies, 70, 147-167.

[114] Tchen, A. H. (1980). "Inequalities for Distributions with Given Marginals." *Annals of Probability* 8, 814-827.

[115] van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

[116] van der Vaart, A. W. and Jon A. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer.

[117] Vijverberg, W. P. M. (1993), "Measuring the Unidentified Parameter of the Extended Roy Model of Selectivity," *Journal of Econometrics* 57, 69-89.

[118] Williamson, R. C. and T. Downs (1990). "Probabilistic Arithmetic I: Numerical Methods for Calculating Convolutions and Dependency Bounds." *International Journal of Approximate Reasoning* 4, 89-158.

[119] Whittington, E. H., H. P. Bain, and C. M. Achilles. (1985) "Effects of Class Size on First-Grade Students." *Spectrum, Journal of School Research and Information* 3, 33-39.

[120] Word, E. Johnson, J, Bain, H. P., Fulton, B. D., Zaharias, J. B., Achilles C. M., Lintz, M. N., Folger, J., and Breda, C. (1990a) *The State of Tennessee's Student/Teacher Achievement Ratio (STAR) Project Final Summary Report.*

[121] ___________________ (1990b) *The State of Tennessee's Student/Teacher Achievement Ratio (STAR) Project Technical Report 1985-1990.*

[122] Wu, C. F. J. (1990). "On the Asymptotic Properties of the Jackknife Histogram." *Annals of Statistics* 18, 1438-1452.