Confidence Sets for Some Partially Identified Parameters*

Yanqin Fan and Sangsoo Park
Department of Economics
Vanderbilt University
VU Station B #351819
2301 Vanderbilt Place
Nashville, TN 37235-1819

This version: October 2007

Abstract

In this paper, we re-visit the inference problem for interval identified parameters originally studied in Imbens and Manski (2004) and later extended in Stoye (2007). We establish a new confidence interval that is asymptotically valid under the same assumptions as in Stoye (2007). Like the confidence interval of Stoye (2007), our new confidence interval extends that of Imbens and Manski (2004) to allow for the lack of a super-efficient estimator of the length of the identified interval. In addition, it shares the natural nesting property of the original confidence interval of Imbens and Manski (2004). A simulation study is conducted to examine the finite sample performance of our new confidence interval and that of Stoye (2007). Finally we extend our CI for interval identified parameters to parameters defined by moment equalities/inequalities.

*We thank Patrik Guggenberger, Chuck Manski, Frank Schorfheide, Kevin Song, Jörg Stoye, and Jisong Wu for helpful discussions. We are espsecially grateful to Patrik Guggenberger, Frank Schorfheide, and Jörg Stoye for sending us their papers, and to Francesco Molinari for providing the data set used in the simulation study in this paper.

1 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 Fan and Park (2007a, b) 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 paper 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 paper, 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 paper 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.

2 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 $\widehat{\theta}_l, \widehat{\theta}_u$ and other assumptions, including super-efficiency of the estimator of $\Delta \equiv \theta_u - \theta_l$, where $\widehat{\theta}_l, \widehat{\theta}_u$ 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 Fan and Park (2007a,b), 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 Fan and Park (2007a). 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:

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

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:

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

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.

2.1 A Review of IM and Stoye (2007)

IM proposed a CI for θ_0 as follows:

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

where c_α solves

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

in which $\widehat{\Delta} = \widehat{\theta}_u - \widehat{\theta}_l$ and $\widehat{\theta}_l, \widehat{\theta}_u, \widehat{\sigma}_l, \widehat{\sigma}_u$ 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 $\widehat{\theta}_l, \widehat{\theta}_u$ that satisfy

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

uniformly in $P \in \mathcal{P}$, and there are estimators $\left(\widehat{\sigma}_l^2, \widehat{\sigma}_u^2, \widehat{\rho}\right)$ 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 $\lim_{n\to\infty} \inf_{\theta\in\Theta} \inf_{P:\theta_0(P)=\theta} P\left(\theta_0 \in CI_{\mathrm{IM}}\right) = 1 - \alpha$, 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:

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

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

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

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

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

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

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

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 $\widehat{\Delta}$.

2.2 A 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 $\left(\widehat{\theta}_l, \widehat{\theta}_u\right)$, i.e., Assumption IM (i), and the specific structure of the identified set $[\theta_l, \theta_u]$ through the use of $\widehat{\Delta}$ or Δ^*, see e.g., (3) and (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 $\left(\widehat{\theta}_l, \widehat{\theta}_u\right)$ 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 $\left(\widehat{\theta}_l, \widehat{\theta}_u\right)$ 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:

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

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

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

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 $(Z_{l,h_\rho} - h_l)_+^2 + (Z_{u,h_\rho} + h_u)_-^2$, 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 $cv_{1-\alpha}(h_l, h_u, h_\rho)$ 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 $c_{1-\alpha}(\theta)$ in CS_n as $cv_{1-\alpha}(0, 0, \widehat{\rho})$. This is the analog of PA-CS introduced in AG (2007) for parameters defined by moment equalities/inequalities. Specifically,

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

We show in Appendix B that CI_{AG} is in fact an interval, since $cv_{1-\alpha}(0, 0, \widehat{\rho})$ 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 $\sigma_u h_u + \sigma_l h_l = \lim(\sqrt{n}\Delta)$. In the special case where $\widehat{\rho} = 1$, $J_{(0,0,1)}$ is $\chi_{[1]}^2$ 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):

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

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

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

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

$$CI_{\text{FP}} = \{\theta : T_n(\theta) \leq c_{1-\alpha}^*(\widehat{\rho})\},$$

where

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

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

THEOREM 2.1 *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 $c^*_{1-\alpha}(\widehat{\rho})$ 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]

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

or equivalently

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

It follows from (8) and the form of CI_{FP} established in Appendix C that with super-efficiency or $\widehat{\rho} = 1$, CI_{FP} reduces to the uniform CI for θ_0 proposed in IM except that CI_{FP} uses Δ^*, while IM uses $\widehat{\Delta}$. In this sense, CI_{FP} can be regarded as a natural extension of 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 $c^*_{1-\alpha}(\widehat{\rho})$ defined in (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 Fan and Park (2007b) provides an example of interval identified parameters for which the asymptotic distribution of estimators of the sharp bounds exhibits discontinuity as a function of paramerers 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.

2.3 The 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

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

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

Then a CI for θ_0 can be defined as

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

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

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

We obtain

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

where

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

The second equality follows from concavity of $\Pr\left(Z_{l,h_\rho} \le h_l + \sqrt{c_l} \cap Z_{u,h_\rho} \ge -h_u - \sqrt{c_u}\right)$ 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

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

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

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

3 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:

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

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 (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 (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 $\pi_j(Y_j, Y_{-j}) > 0$; 0 otherwise. Let's assume a simple linear payoff function, that is, $\pi_j(Y_j, Y_{-j}) = X_j\beta_j - d_jY_{-j} + v_j$, $E[v_j|X_j, X_{-j}] = 0$, and $d_j > 0$. Then, because there exist multiple equilibria, $E[Y_1(1-Y_0)|X_1, X_2] = P(Y_1 = 1, Y_0 = 0|X_1, X_2)$ is partially identified i.e.

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

where

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

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

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.

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

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 $P(Y_{Li} \leq Y_i \leq Y_{Ui}) = 1$. The parameter θ satisfies

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

These conditional restrictions imply the inequalities

$$E[Y_{Li}Z_i] \leq \theta' E[X_i Z_i] \leq E[Y_{Ui}Z_i], \tag{12}$$

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 (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 (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 (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 (11) as moment equalities:

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

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

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

where $\widehat{\sigma}_{n,j}^2(\theta)$ 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

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

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

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

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 $cv_{1-\alpha}\left(0, \widehat{\Omega}_n(\theta)\right)$. 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

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

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

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

THEOREM 3.1 *Under the same assumptions as Theorem 2 (a) of AG (2007), we have*

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

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 $cv_{1-\alpha}\left(h_1, \widehat{\Omega}_n(\theta)\right)$ except that they use different values of h_1: The CS of PPHI (2006) and the PA-CS of AG (2007) use $cv_{1-\alpha}\left(0, \widehat{\Omega}_n(\theta)\right)$ 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 $cv_{1-\alpha}\left(0, .., 0, \infty, .., \infty, \widehat{\Omega}_n(\theta)\right)$ 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:

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

An alternative CS for θ_0 can be defined as follows:

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

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, $\theta - \widehat{\theta}_l$ 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_{MI}, 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

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

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

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

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

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 $cv_{1-\alpha}(0,0,\widehat{\rho}(\theta))$, where

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

We propose to use:

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

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

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

4 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}.

4.1 Computation and Comparison of Critical Values

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

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

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

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

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

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

implying (14).

So to compute $c^*_{1-\alpha}(\rho)$, we just need to compute $cv_{1-\alpha}\left(\frac{\sqrt{n}\Delta^*}{\widehat{\sigma}_l}, 0, \widehat{\rho}\right)$ or $cv_{1-\alpha}\left(0, \frac{\sqrt{n}\Delta^*}{\widehat{\sigma}_u}, \widehat{\rho}\right)$ depending on which of $\widehat{\sigma}_l, \widehat{\sigma}_u$ 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

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

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 $c^*_{1-\alpha}(\widehat{\rho})$ 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 below, 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 $\rho \to 1$. But for small values of ρ, $cv_{1-\alpha}(0,0,\rho)$ can be much larger than $cv_{1-\alpha}(0,0,1)$.

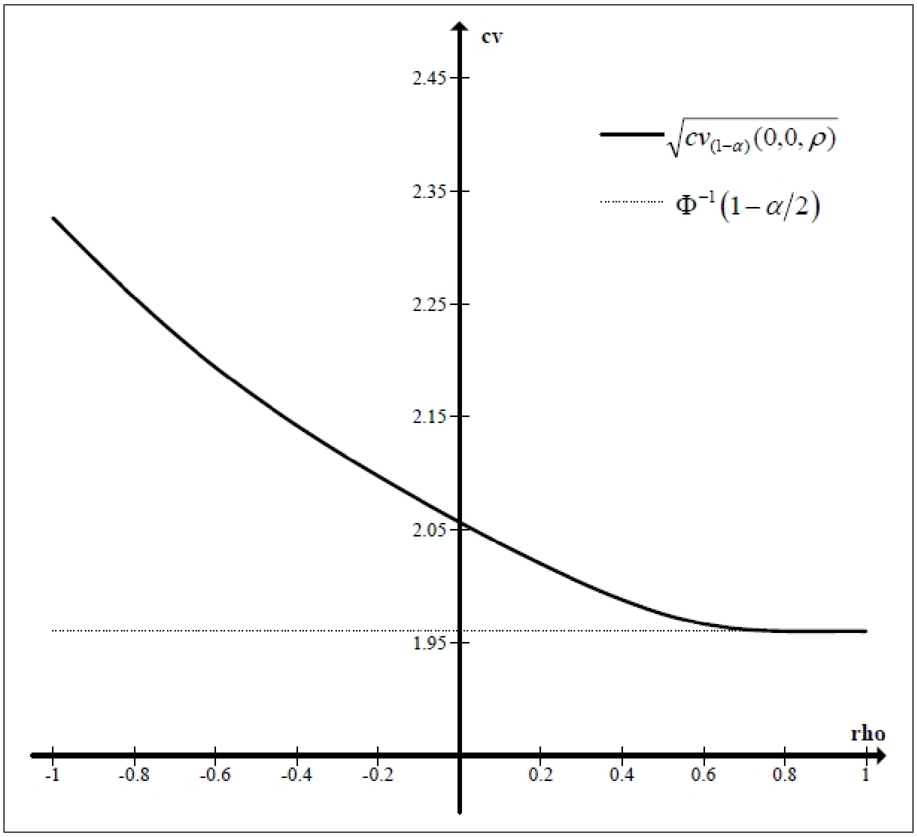


Figure 1. $\sqrt{cv_{0.95}(0,0,\rho)}$ and $\Phi^{-1}(0.975)$

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

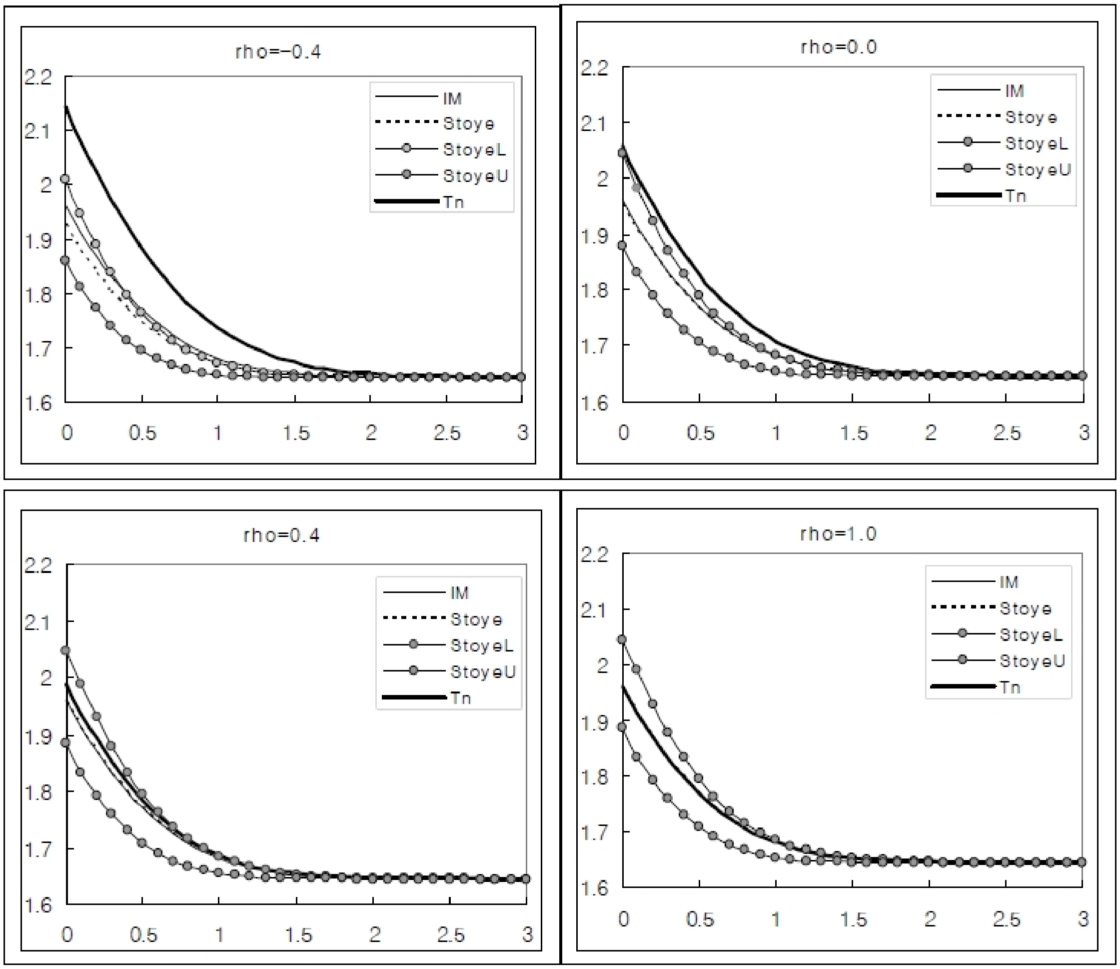


Figure 2. Comparison of critical values

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: $\left(\sigma_l^2, \sigma_u^2\right) = (2, 2)$ and $\left(\sigma_l^2, \sigma_u^2\right) = (1, 2)$. 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 $\sigma_l = \sigma_u$, 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}.

4.2 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 the following table.

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 below.

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 2.1. 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 $\widehat{\theta}_l$ is larger than $\widehat{\theta}_u$ in which case, the effect of using the shrinkage estimator with $b_n = 0$ is to replace negative $\widehat{\Delta}$'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 $\widehat{\Delta}$. When $c = 3.5, 4$, b_n satisfies the conditions of Theorem 2.1, 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 $\theta = \theta_l, \theta_u$.

4.2.1 Point-identified case

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

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

This is also the CI of IM and Horowitz and Manski (2000). Its coverage rates denoted by $\text{CR}(\theta_0)$ and width over 2000 simulations are reported in Table 3 below.

Table 3: Summary Statistics for CI_n

n	$\text{CR}(\theta_0)$	Width
500	0.9485	0.1720
1000	0.9525	0.1219
2000	0.950	0.0861
8000	0.9520	0.0431

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, $\widehat{\Delta}$ may not be exactly zero. In fact, it is possible that $\widehat{\Delta}$ 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 below, 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 $\widehat{\Delta}$. 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].

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

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

[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.

4.2.2 Interval-identified case

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

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**

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 $c = 0$, $P(\Delta^*) = 0$, 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: A 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.

5 Conclusion and Current Research

In this paper, 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 paper support the theoretical finding of Stoye (2007) and the current paper: 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 paper 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 paper is violated. For example, the validity of CI_{FP} relies on the assumption that the asymptotic distribution of $\left(\widehat{\theta}_l, \widehat{\theta}_u\right)$ 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 $\theta_l \le \theta_0 \le \theta_u$. 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.

6 Appendix A: Technical Proofs

Proof of Theorem 2.1. Let

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

and

$$\overline{CI}_{\text{FP}} = \{\theta : T_n(\theta) \le \bar{c}_{1-\alpha}(\rho)\}.$$

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: $\lim_{n\to\infty} \inf_{\theta\in\Theta} \inf_{P:\theta_0(P)=\theta} P\left(\theta_0 \in \overline{CI}_{\text{FP}}\right) = 1-\alpha$. 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 $\bar{c}_{1-\alpha}(\rho)$ is replaced with $c^*_{1-\alpha}(\widehat{\rho})$. Since $\hat{\sigma}_l$, $\hat{\sigma}_u$, and $\widehat{\rho}$ are uniformly consistent estimators of σ_l, σ_u, and ρ respectively, the result holds with

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

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. $\Delta_n \ge c_n$; Case II. $\Delta_n < c_n$.

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

Suppose $h_l = \infty$. Then

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

where we have used the result that $\Pr\left[\Delta^* = \widehat{\Delta}\right] \to 1$ because of $\Pr\left[\widehat{\Delta} > b_n\right] \to 1$. 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 $\Delta^* = 0 \le \Delta$ with probability approaching one. Note that

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

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

Proof of Theorem 3.1. 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

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

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

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

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,

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

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,

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

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

7 Appendix B: An Expression for $J_h(x)$

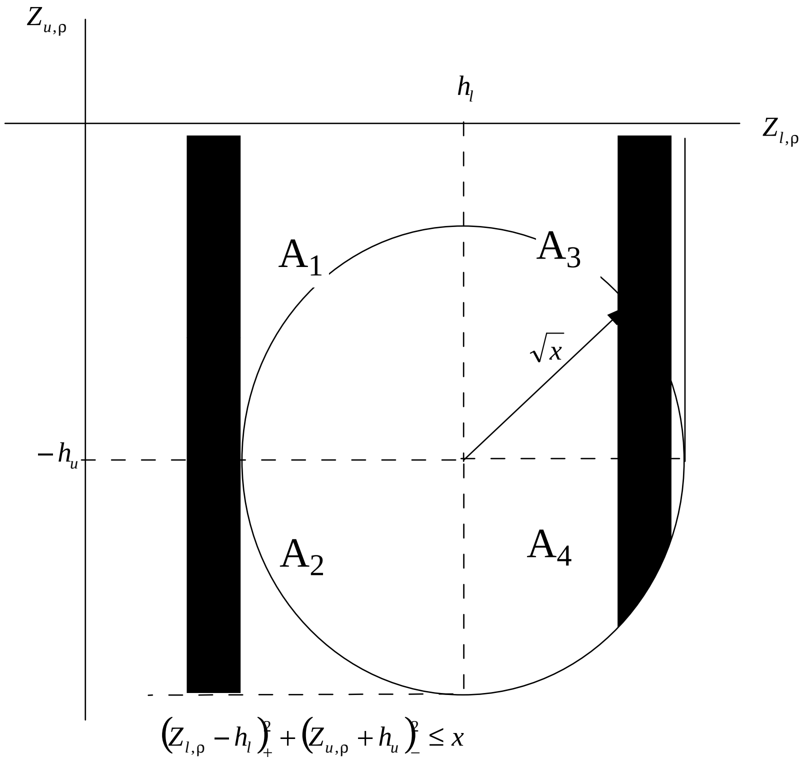
In this section, we derive a closed form expression for $J_h(x)$. This should be useful in constructing CSs in moment inequality models when there are two moment constraints. Let $\phi(z_l, z_u; \rho)$ and $\Phi(z_l, z_u; \rho)$ denote respectively the pdf and cdf of $(Z_{l,\rho}, Z_{u,\rho})$: the standard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

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

If $|\rho| < 1$, then

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

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



Hence,

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

If $\rho = 1$, then

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

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

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

Therefore, we get

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

If $\rho = -1$, then

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

Let $\max\{h_l,h_u\} = h_{\max}$ and $\min\{h_l,h_u\} = h_{\min}$. We can rewrite the event $\left\{(Z-h_l)_+^2 + (Z-h_u)_+^2 \le x\right\}$ as:

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

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,

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

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

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

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

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

Similarly, we can show:

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

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

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

Combining them altogether, we get

$$\begin{aligned}
&\left\{(Z-h_l)_+^2 + (Z-h_u)_+^2 \le x\right\} \\
= &\left(-\infty, \min\left\{h_{\max}, h_{\min}+\sqrt{x}\right\}\right) \bigcup \left\{\begin{array}{c} \varnothing \text{ if } x \le (h_{\max}-h_{\min})^2 \\ \left(h_{\max}, \frac{h_l+h_u+\sqrt{2x-(h_{\max}-h_{\min})^2}}{2}\right] \text{ if o/w} \end{array}\right. \\
= &\left\{\begin{array}{l} \left(-\infty, h_{\min}+\sqrt{x}\right) \text{ if } x \le (h_{\max}-h_{\min})^2 \\ \left(-\infty, \frac{h_l+h_u+\sqrt{2x-(h_{\max}-h_{\min})^2}}{2}\right) \text{ if o/w} \end{array}\right.
\end{aligned}$$

Therefore,

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

8 Appendix C. The Form of the Confidence Set CS_n

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

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

We need to distinguish between two cases. Case I. $\hat{\theta}_l \le \hat{\theta}_u$ and Case II. $\hat{\theta}_l \ge \hat{\theta}_u$. For Case I, it is easy to show that

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

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

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

where

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

By definition, we obtain

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

and

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

Now,

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

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

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

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

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

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

Now, one can show:

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

and

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

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

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

References

[1] 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.

[2] Andrews, D.W.K. and P. Guggenberger (2005b), "Hybrid and Size-corrected Subsampling Methods," Unpublished Manuscript, Cowles Foundation, Yale University.

[3] Andrews, D.W.K. and P. Guggenberger (2005c), "Applications of Subsampling, Hybrid, and Size-correction Methods," Cowles Foundation Discussion Paper No. 1608, Yale University.

[4] 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.

[5] Beresteanu, A. and F. Molinari (2006), "Asymptotic Properties for a Class of Partially Identified Models," Manuscript.

[6] Berry, S.T. (1992), "Estimation of a Model of Entry in the Airline Industry", *Econometrica* 60, 889-917.

[7] Bresnahan, T. and P. Reiss (1991), "Empirical models of discrete games," *Journal of Econometrics* 48, 57–82.

[8] Bugni, F. A. (2007), "Bootstrap Inference in Partially Identified Models," mimeo, Northwestern University.

[9] Chernozhukov, V., H. Hong, and E. Tamer (2004), "Inference on Parameter Sets in Econometric Models," Working Paper, MIT.

[10] Chernozhukov, V., H. Hong, and E. Tamer (2007), "Parameter Set Inference in a Class of Econometric Models," *Econometrica* 75, 1243-1284.

[11] Ciliberto, F. and E. Tamer (2004), “Market Structure and Multiple Equilibria in Airline Markets,” mimeo, Northwestern University.

[12] Ding, W. and S. Lehrer (2005), "Class Size and Student Achievement: Experimental Estimates of Who Benefits and Who Loses from Reductions", *Queen's Economics Department Working Paper* 1046, Queen's University.

[13] Fan, Y. and S. Park (2007a), “Statistical Inference on the Quantile of the Treatment effects,” in progress.

[14] Fan, Y. and S. Park (2007b), “Sharp Bounds on the Distribution of the Treatment effects and Their Statistical Inference,” Unpublished Manuscript, Vanderbilt University.

[15] Fan, Y. and J. Wu (2007), “Sharp Bounds on the Distribution of the Treatment effects in Switching Regimes Models,” Unpublished Manuscript, Vanderbilt University.

[16] Horowitz, J. and C.F. Manski (2000), “Nonparametric Analysis of Randomized Experiments with Missing Covariates and Outcome Data,” *Journal of the American Statistical Association* 95, 77-84.

[17] Imbens, G. W. and C. F. Manski (2004), “Confidence Intervals For Partially Identified Parameters,” *Econometrica* 72, 1845–1857.

[18] Manski, C.F. (1997), “The Mixing Problem in Program Evaluation,” *Review of Economic Studies* 64, 537-553.

[19] Manski, C. F. (2003), *Partial Identification of Probability Distributions.* Springer.

[20] Manski, C.F. and E. Tamer (2002), “Inference on Regressions with Interval Data on a Regressor or Outcome,” *Econometrica* 70, 519-546.

[21] Moon, R. and F. Schorfheide (2007), “A Bayesian Look at Partially-Identified Models,” Manuscript, U.Penn.

[22] Pakes, A., J. Porter, K. Ho, and J. Ishii (2006), “Moment Inequalities and Their Application,” Working Paper, ???

[23] Park, S. (2007a), “Confidence Intervals for the Distribution of the Treatment effects,” in progress.

[24] Park, S. (2007b), “Heterogeneous Effects of Class Size Reduction: Re-visit the Project STAR,” in progress.

[25] Park, S. (2007c), “The Mixing Problem in Program Evaluation: Inference,” in progress.

[26] 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.

[27] 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.

[28] Rosen, A.M. (2005), "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," mimeo, University College London.

[29] Soares, G. (2006), "Inference for Partially Identified Models with Inequality Moment Constraints," Working Paper, Yale.

[30] Stoye, J. (2007), "More on Confidence Intervals for Partially Identified Parameters," Working paper, NYU.

[31] Vijverberg, W. P. M. (1993), "Measuring the Unidentified Parameter of the Extended Roy Model of Selectivity," *Journal of Econometrics* 57, 69-89.

[32] Tamer, E. (2003), "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," Review of Economic Studies, 70, 147-167.

[33] 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.