

# Confidence Sets for Some Partially Identified Parameters\*

Yanqin Fan<sup>†</sup> and Sang Soo Park<sup>‡</sup>

First version: November 2007

This version: May 2009

## Abstract

In this paper, we first re-visit the inference problem for interval identified parameters originally studied in Imbens and Manski (2004) and later extended in Stoye (2007). We take the general criterion function approach and 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 confidence interval for interval identified parameters to parameters defined by moment equalities/inequalities.

*Keywords:* Interval identified parameters; Moment equalities/inequalities; Criterion function approach.

*JEL Classification:* C12, C13, C14, C19

---

\*We thank Patrik Guggenberger, Chuck Manski, Frank Schorfheide, Kevin Song, Jörg Stoye, and Jisong Wu for helpful discussions. We are especially grateful to Patrik Guggenberger, Frank Schorfheide, and Jörg Stoye for sending us their papers, to Gustavo Soares for sending us his dissertation, and to Francesca Molinari for providing the data set used in the simulation study in this paper. **Sang Soo Park specially thanks the attenders at his job seminars at the Departments of Economics in University of North Carolina, University of New South Wales, Bocconi University, Ohio University, IZA, and KIPF for their valuable comments and remarks.**

<sup>†</sup>Department of Economics, Vanderbilt University, VU Station B #351819, 2301 Vanderbilt Place, Nashville, TN 37235-1819, USA.

<sup>‡</sup>Department of Economics, University of North Carolina, CB #3305 Gardner Hall, Chapel Hill, NC 27599-3305, USA.

# 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), Park (2007a) 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, including Bugni (2007), Canay (2007), CHT, Galichon and Henry (2006), Romano and Shaikh (2005a,b), Stoye (2007), Rosen (2005), Soares (2006), Beresteanu and Molinari (2006), Moon and Schorfheide (2007), and Andrews and Guggenberger (2007) (AG (2007) henceforth), among others.

The simplicity of the CIs of IM 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 based on the duality between CSs and hypotheses tests. CHT first applied this idea to constructing CSs for partially identified parameters in a very general set-up and referred to it as the criterion function approach. In AG (2007), they proposed a simple plug-in asymptotic CS (PA-CS) for parameters defined by moment equalities/inequalities and 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.

The current paper builds on the work of IM, Stoye (2007), AG (2007), and CHT. It makes several contributions to the literature on inference for partially identified parameters. First, we construct an asymptotically uniformly valid, non-conservative CS for interval identified parameters originally studied in IM by using the general criterion function approach, see CHT, and show that it reduces to the CI of IM when there exists a super-efficient estimator of the length of the identified interval. Given that IM is the first to propose and develop CIs for partially (interval) identified parameters

and the CIs of IM have been employed in several empirical work, we feel it's important to understand its relation with the CSs established by the general criterion function approach for interval identified parameters. Second, we show that for interval identified parameters, the PA-CS and the CS we construct by using the general criterion function approach are in fact CIs, although the form of the CI depends on how close the estimators of the lower and upper bound of the identified interval are to each other. We establish the closed-form expressions for the different intervals that constitute our CS and the PA-CS. These closed-form expressions greatly simplify the computation of the CS, as in general, one needs to check if each and every parameter value in the parameter space is in the CS which can be extremely time consuming. Third, we show that the CI of Stoye (2007) can be obtained by inverting two one-sided tests for the true parameter. 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. Fourth, we extend the CI of IM to CSs for parameters defined by general moment equalities/inequalities. 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 corresponding 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.

We carried out a simulation study on interval data and applied our new confidence interval, that of Stoye (2007), and the PA-CS of CHT and AG (2007) to three artificially created data generating processes (DGP) from the March 2000 wave of the Current Population Survey (CPS). 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 can over-cover when the length of the identified interval is bounded away from zero 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 hy-

pothesis 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, including the closed-form expressions for the CS we develop and the PA-CS for interval identified parameters.

## 2 Confidence Intervals for Interval-Identified Parameters

Let  $\theta_l \leq \theta_0 \leq \theta_u$ , where  $\theta_0 = \theta_0(P)$  is the parameter 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  $\theta_l, \theta_u$  are identified, but  $\theta_0$  may not be. IM first introduced a uniform CI for  $\theta_0$  under the assumption of asymptotic joint normality of  $\hat{\theta}_l, \hat{\theta}_u$  and other assumptions, including super-efficiency of the estimator of  $\Delta \equiv \theta_u - \theta_l$ , where  $\hat{\theta}_l, \hat{\theta}_u$  are consistent estimators of  $\theta_l, \theta_u$  respectively. Stoye (2007) proposed a uniform CI that does not depend on the super-efficiency condition.

In addition to 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), Park (2007a), 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  $\theta_l$  and  $\theta_u$  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  $\hat{\theta}_u = n^{-1} \sum_{i=1}^n Y_{Ui}$ .

**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  $\Delta$  can be found in Williamson and Downs (1990), see also Fan and Park (2007a). Specifically, for  $0 < p < 1$ , let  $\theta_0 = F_\Delta^{-1}(p)$ ,

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

Then  $\theta_l \leq \theta_0 \leq \theta_u$ . With randomized data,  $F_1$  and  $F_0$  are identified and thus  $\theta_l, \theta_u$  are identified. Estimators of  $\theta_l, \theta_u$  can be constructed by replacing  $F_1$  and  $F_0$  in the above expressions with their

consistent estimators such as the empirical distributions.

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

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

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 a 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 otherwise,  $Y_{1i}$  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,  $\rho_{10}$  is not identified. Using the fact that the covariance matrix of the errors is positive semi-definite, Vijverberg (1993) showed that  $\rho_L \leq \rho_{10} \leq \rho_U$ , where

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

Note that  $\rho_L$  and  $\rho_U$  depend on the identified parameters only and hence are themselves identified, but  $\rho_{10}$  is only interval identified unless  $\rho_L = \rho_U$ . Estimators of  $\rho_L, \rho_U$  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  $\theta_0$  as follows:

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

where  $c_\alpha$  solves

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

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  $CI_{\text{IM}}$ .

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

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

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

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

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

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  $\Delta_n \rightarrow 0$  only if  $\sigma_l^2 - \sigma_u^2 \rightarrow 0$  and  $\rho \rightarrow 1$  for all those sequences. To relax Assumption IM (iii), Stoye (2007) proposed the following CI for  $\theta_0$  and verified its asymptotic uniform validity and non-conservativeness under Assumption IM (i) and (ii) only:

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

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

$$\begin{aligned} \Pr\left(-c_l \leq Z_1 \wedge \widehat{\rho} Z_1 \leq c_u + \frac{\sqrt{n}\Delta^*}{\widehat{\sigma}_u} + \sqrt{1 - \widehat{\rho}^2} Z_2\right) &\geq 1 - \alpha, \\ \Pr\left(-c_l - \frac{\sqrt{n}\Delta^*}{\widehat{\sigma}_l} + \sqrt{1 - \widehat{\rho}^2} Z_2 \leq \widehat{\rho} Z_1 \wedge Z_1 \leq c_u\right) &\geq 1 - \alpha, \end{aligned} \quad (4)$$

in which  $Z_1$  and  $Z_2$  are independent standard normal random variables, and  $\Delta^*$  is a shrinkage estimator of  $\Delta$  defined as

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

and  $b_n$  is some pre-assigned sequence such that  $b_n \rightarrow 0$  and  $b_n\sqrt{n} \rightarrow \infty$ . As shown in Stoye (2007), if Assumption IM (iii) holds, then  $CI_S$  reduces to that of IM (2004) except that  $CI_S$  uses  $\Delta^*$  and  $CI_{IM}$  uses  $\widehat{\Delta}$ . **As emphasized in Stoye (2007), the CI of Stoye is empty, i.e.,  $CI_S = \emptyset$  if  $\widehat{\theta}_l$  is far larger than  $\widehat{\theta}_u$  so that  $\widehat{\theta}_l - \frac{c_l\widehat{\sigma}_l}{\sqrt{n}} > \widehat{\theta}_u + \frac{c_u\widehat{\sigma}_u}{\sqrt{n}}$  or  $\sqrt{n}\widehat{\Delta} < -(c_l\widehat{\sigma}_l + c_u\widehat{\sigma}_u) < 0$ .**

## 2.2 A New Confidence Interval for $\theta_0$

The CIs of IM and Stoye (2007) are computationally simple, but they rely heavily on the asymptotic normality of  $(\widehat{\theta}_l, \widehat{\theta}_u)$ , i.e., Assumption IM (i), and the specific structure of the identified set  $[\theta_l, \theta_u]$  through the use of  $\widehat{\Delta}$  or  $\Delta^*$ , 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 a simple interval form.

In this subsection, we re-visit the issue of constructing CIs for interval identified parameter  $\theta_0$  by using the general approach of inverting a hypothesis test, aiming at understanding the roles played by the asymptotic normality of  $(\widehat{\theta}_l, \widehat{\theta}_u)$  and the estimator of the length of the identified interval. By taking into account the interval structure of the identified set for  $\theta_0$ , 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 that of IM when superefficiency holds. In addition, 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  $(\widehat{\theta}_l, \widehat{\theta}_u)$  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  $\gamma_{1u} = (\theta_u - \theta)/\sigma_u$ ,  $\gamma_2 = (\theta, \rho)$ ,  $\gamma_3$  denotes the remaining parameters in  $P$ . The parameter space is

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

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  $\theta_0$ :

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

A  $1-\alpha$  CS for  $\theta_0$  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  $\Gamma$  for which  $\omega_n^{1/2} \gamma_{\omega_n, h, 1} \rightarrow h_1 \equiv (h_l, h_u)$ ,  $\gamma_{\omega_n, h, 2} \rightarrow h_2 \equiv (h_\theta, h_\rho)$ . Define

$$H = \{(h_1, h_2) \in R_{+\infty}^2 \times R \times [0, 1] : \exists \text{ a subsequence } \{\omega_n\} \text{ of } \{n\} \text{ and a sequence } \{\gamma_{\omega_n, h} : n \geq 1\}\}.$$

Let  $h = (h_1, h_2)$  and  $J_h$  denote the limiting distribution of  $T_n(\theta)$  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  $h_\rho$ , 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  $\theta_0$  using  $J_h$  corresponding to different values of  $h$ . The first one defines the critical value  $c_{1-\alpha}(\theta)$  in  $CS_n$  as  $cv_{1-\alpha}(0, 0, \hat{\rho})$ . This is the analog of PA-CS introduced in AG (2007) for parameters defined by moment equalities/inequalities, see also CHT. Specifically,

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

We show in Appendix C that  $CI_{\text{PA}}$  is in fact an interval, since  $cv_{1-\alpha}(0, 0, \hat{\rho})$  does not depend on  $\theta$ . Note that  $h_l \geq 0$ ,  $h_u \geq 0$ , and  $J_h$  is stochastically decreasing in  $h_l, h_u$ , implying

$$cv_{1-\alpha}(0, 0, h_\rho) = \sup_{h_l \geq 0, h_u \geq 0} cv_{1-\alpha}(h_l, h_u, h_\rho).$$

Since  $h_\rho$  can be consistently estimated by  $\hat{\rho}$ , it follows that  $CI_{\text{PA}}$  is asymptotically uniformly valid, but it is conservative when  $\Delta$  is bounded away from zero or when  $\Delta$  is a known but non-zero constant. The reason for the latter is that  $(0, 0, h_\theta, h_\rho)$  may not belong to  $H$  unless  $\theta_l = \theta_u$ , as  $h_l, h_u$  satisfy  $\sigma_u h_u + \sigma_l h_l = \lim(\sqrt{n}\Delta)$ . In the special case where  $\hat{\rho} = 1$ ,  $J_{(0,0,1)}$  is  $\chi_{[1]}^2$  and  $CI_{\text{PA}}$  reduces to the symmetric CI for the identification region  $[\theta_l, \theta_u]$  first proposed in Horowitz and Manski (2000):

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

see also (2) in IM, where  $z_\alpha$  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

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

where

$$c_{1-\alpha}^*(\hat{\rho}) = \sup_{h_l \geq 0, h_u \geq 0, \hat{\sigma}_u h_u + \hat{\sigma}_l h_l = \sqrt{n}\Delta^*} cv_{1-\alpha}(h_l, h_u, \hat{\rho}) \quad (7)$$

in which  $\Delta^*$  is the shrinkage estimator defined in (5). We show in Appendix A that  $CI_{FP}$  is asymptotically uniformly valid and non-conservative.

**THEOREM 2.1** *Suppose Assumption IM (i) and (ii) hold and  $0 < \alpha < 1/2$ . Then  $CI_{FP}$  satisfies  $\lim_{n \rightarrow \infty} \inf_{\theta \in \Theta} \inf_{P: \theta_0(P) = \theta} \Pr(\theta_0 \in CI_{FP}) = 1 - \alpha$ .*

We now show that in fact  $c_{1-\alpha}^*(\hat{\rho})$  can be computed easily without any optimization involved. Define

$$\begin{aligned} W(h_l) &\equiv (Z_{l,\hat{\rho}} - h_l)_+^2 + (Z_{u,\hat{\rho}} + h_l)_-^2 \\ &= (Z_{l,\hat{\rho}} - h_l)_+^2 + \left( Z_{u,\hat{\rho}} + \frac{\sqrt{n}\Delta^*}{\hat{\sigma}_u} - \frac{\hat{\sigma}_l}{\hat{\sigma}_u} h_l \right)_-^2. \end{aligned}$$

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

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

i.e.,

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

From the symmetry of the joint distribution of  $(Z_{l,\hat{\rho}}, Z_{u,\hat{\rho}})$ , it follows that the random variable  $(Z_{l,\hat{\rho}})_+^2 + \left( Z_{u,\hat{\rho}} + \frac{\sqrt{n}\Delta^*}{\hat{\sigma}_u} \right)_-^2$  has the same distribution function as the random variable  $\left( Z_{l,\hat{\rho}} - \frac{\sqrt{n}\Delta^*}{\hat{\sigma}_u} \right)_+^2 + (Z_{u,\hat{\rho}})_-^2$ . Thus,  $cv_{1-\alpha} \left( 0, \frac{\sqrt{n}\Delta^*}{\hat{\sigma}_u}, \hat{\rho} \right) = cv_{1-\alpha} \left( \frac{\sqrt{n}\Delta^*}{\hat{\sigma}_u}, 0, \hat{\rho} \right)$ . But since  $\left( Z_{l,\hat{\rho}} - \frac{\sqrt{n}\Delta^*}{\hat{\sigma}_l} \right)_+^2 + (Z_{u,\hat{\rho}})_-^2$  is stochastically increasing in  $\hat{\sigma}_l$ , we have

$$c_{1-\alpha}^*(\hat{\rho}) = cv_{1-\alpha} \left( \frac{\sqrt{n}\Delta^*}{\max\{\hat{\sigma}_l, \hat{\sigma}_u\}}, 0, \hat{\rho} \right). \quad (8)$$

The expression in (8) greatly simplifies the computation of  $c_{1-\alpha}^*(\hat{\rho})$ , in particular, no optimization is needed. One method for computing  $c_{1-\alpha}^*(\hat{\rho})$  is by simulation. Alternatively, one can invert  $J_h$  numerically. In Appendix B, we show that for  $|\rho| < 1$ ,

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

for  $\rho = 1$ ,

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

for  $\rho = -1$ ,

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

where  $h_{\max} = \max\{h_l, h_u\}$  and  $h_{\min} = \min\{h_l, h_u\}$ . For any fixed  $x$ , the value of  $J_h(x)$  can be computed numerically using the above expressions. We have written a Gauss program for computing  $c_{1-\alpha}^*(\hat{\rho})$  which is available upon request.

Similar to  $CI_{PA}$ ,  $CI_{FP}$  is an interval, as  $c_{1-\alpha}^*(\hat{\rho})$  does not depend on  $\theta$ . Interestingly, if  $\rho = 1$ , then  $c_{1-\alpha}^*(1)$  is the  $(1 - \alpha)$  quantile of the distribution  $\Phi\left(\frac{\sqrt{n}\Delta^*}{\max\{\hat{\sigma}_l, \hat{\sigma}_u\}} + \sqrt{x}\right) - \Phi(-\sqrt{x})$  and thus satisfies<sup>1</sup>

$$\Phi\left(\frac{\sqrt{n}\Delta^*}{\max\{\hat{\sigma}_l, \hat{\sigma}_u\}} + \sqrt{c_{1-\alpha}^*(1)}\right) - \Phi\left(-\sqrt{c_{1-\alpha}^*(1)}\right) = 1 - \alpha. \quad (9)$$

It follows from (9) and the form of  $CI_{FP}$  established in Appendix C that when  $\hat{\rho} = 1$ ,  $CI_{FP}$  reduces to the uniform CI for  $\theta_0$  proposed in IM except that  $CI_{FP}$  uses  $\Delta^*$ , while IM uses  $\hat{\Delta}$ . In this sense,  $CI_{FP}$  can be regarded as a natural extension of IM from  $\rho = 1$  to any  $\rho$ .

**The explicit expressions for  $CI_{FP}$  are the following.** Let  $\hat{\sigma}_{\min} = \min\{\hat{\sigma}_l, \hat{\sigma}_u\}$  and  $\hat{\sigma}_{\max} = \max\{\hat{\sigma}_l, \hat{\sigma}_u\}$ .

$$CI_{FP} = \begin{cases} \left[ \hat{\theta}_l - \sqrt{c_{1-\alpha}^*(\hat{\rho})} \frac{\hat{\sigma}_l}{\sqrt{n}}, \hat{\theta}_u + \sqrt{c_{1-\alpha}^*(\hat{\rho})} \frac{\hat{\sigma}_u}{\sqrt{n}} \right] & \text{if } \hat{\Delta} \geq -\sqrt{c_{1-\alpha}^*(\hat{\rho})} \frac{\hat{\sigma}_{\min}}{\sqrt{n}} \\ \left[ \hat{\theta}_l - \sqrt{c_{1-\alpha}^*(\hat{\rho})} \frac{\hat{\sigma}_l}{\sqrt{n}}, B \right] & \text{if } -\sqrt{c_{1-\alpha}^*(\hat{\rho})} \frac{\hat{\sigma}_l}{\sqrt{n}} \leq \hat{\Delta} < \sqrt{c_{1-\alpha}^*(\hat{\rho})} \frac{\hat{\sigma}_u}{\sqrt{n}} \\ \left[ A, \hat{\theta}_u + \sqrt{c_{1-\alpha}^*(\hat{\rho})} \frac{\hat{\sigma}_u}{\sqrt{n}} \right] & \text{if } -\sqrt{c_{1-\alpha}^*(\hat{\rho})} \frac{\hat{\sigma}_u}{\sqrt{n}} \leq \hat{\Delta} < \sqrt{c_{1-\alpha}^*(\hat{\rho})} \frac{\hat{\sigma}_l}{\sqrt{n}} \\ [A, B] & \text{if } -\sqrt{c_{1-\alpha}^*(\hat{\rho})} \frac{\sqrt{\hat{\sigma}_u^2 + \hat{\sigma}_l^2}}{\sqrt{n}} \leq \hat{\Delta} < -\sqrt{c_{1-\alpha}^*(\hat{\rho})} \frac{\hat{\sigma}_{\max}}{\sqrt{n}} \\ \emptyset & \text{if } \hat{\Delta} < -\sqrt{c_{1-\alpha}^*(\hat{\rho})} \frac{\sqrt{\hat{\sigma}_u^2 + \hat{\sigma}_l^2}}{\sqrt{n}} \end{cases}, \quad (10)$$

where

$$A \equiv \frac{\hat{\sigma}_u^2 \hat{\theta}_l + \hat{\sigma}_l^2 \hat{\theta}_u}{\hat{\sigma}_u^2 + \hat{\sigma}_l^2} - \sqrt{\frac{\hat{\sigma}_l^2 \hat{\sigma}_u^2}{n(\hat{\sigma}_u^2 + \hat{\sigma}_l^2)} \left[ c_{1-\alpha}^*(\hat{\rho}) - \frac{n\hat{\Delta}^2}{(\hat{\sigma}_u^2 + \hat{\sigma}_l^2)} \right]},$$

$$B \equiv \frac{\hat{\sigma}_u^2 \hat{\theta}_l + \hat{\sigma}_l^2 \hat{\theta}_u}{\hat{\sigma}_u^2 + \hat{\sigma}_l^2} + \sqrt{\frac{\hat{\sigma}_l^2 \hat{\sigma}_u^2}{n(\hat{\sigma}_u^2 + \hat{\sigma}_l^2)} \left[ c_{1-\alpha}^*(\hat{\rho}) - \frac{n\hat{\Delta}^2}{(\hat{\sigma}_u^2 + \hat{\sigma}_l^2)} \right]}.$$

<sup>1</sup>As explicitly stated in (9), the critical values for IM in (3) are comparable with  $\sqrt{c_{1-\alpha}^*(1)}$  instead of  $c_{1-\alpha}^*(1)$ , due to the different ways in which  $CI_{FP}$  and  $CI_{IM}$  are expressed.

Consider a simple case that  $\hat{\sigma}_l = \hat{\sigma}_u$  in order to understand the implication of  $CI_{\text{FP}}$ .

When  $\hat{\sigma}_l = \hat{\sigma}_u = \hat{\sigma}$ ,

$$CI_{\text{FP}} = \begin{cases} \left[ \hat{\theta}_l - \sqrt{c_{1-\alpha}^*(\hat{\rho})} \frac{\hat{\sigma}}{\sqrt{n}}, \hat{\theta}_u + \sqrt{c_{1-\alpha}^*(\hat{\rho})} \frac{\hat{\sigma}}{\sqrt{n}} \right] & \text{if } \hat{\Delta} \geq -\sqrt{c_{1-\alpha}^*(\hat{\rho})} \frac{\hat{\sigma}}{\sqrt{n}} \\ [A, B] & \text{if } -\sqrt{2c_{1-\alpha}^*(\hat{\rho})} \frac{\hat{\sigma}}{\sqrt{n}} \leq \hat{\Delta} < -\sqrt{c_{1-\alpha}^*(\hat{\rho})} \frac{\hat{\sigma}}{\sqrt{n}} \\ \emptyset & \text{if } \hat{\Delta} < -\sqrt{2c_{1-\alpha}^*(\hat{\rho})} \frac{\hat{\sigma}}{\sqrt{n}} \end{cases} \quad (11)$$

and  $A$  and  $B$  become

$$A \equiv \frac{\hat{\theta}_l + \hat{\theta}_u}{2} - \frac{\hat{\sigma}}{2\sqrt{n}} \sqrt{c_{1-\alpha}^*(\hat{\rho}) - \frac{n\hat{\Delta}^2}{2\hat{\sigma}^2}}, \quad B \equiv \frac{\hat{\theta}_l + \hat{\theta}_u}{2} + \frac{\hat{\sigma}}{2\sqrt{n}} \sqrt{c_{1-\alpha}^*(\hat{\rho}) - \frac{n\hat{\Delta}^2}{2\hat{\sigma}^2}}.$$

Similarly to the CI of Stoye (2007), the  $CI_{\text{FP}}$  is empty when  $\hat{\theta}_l$  is too far above  $\hat{\theta}_u$  such that  $\sqrt{n}\hat{\Delta}/\hat{\sigma} < -\sqrt{2c_{1-\alpha}^*(\hat{\rho})}$  and it takes the standard form if  $\sqrt{n}\hat{\Delta}/\hat{\sigma} \geq -\sqrt{c_{1-\alpha}^*(\hat{\rho})}$ . But interestingly, there is a middle case where  $\hat{\theta}_l$  is larger than  $\hat{\theta}_u$ , but  $\hat{\Delta}$  satisfies  $-\sqrt{2c_{1-\alpha}^*(\hat{\rho})} \leq \sqrt{n}\hat{\Delta}/\hat{\sigma} < -\sqrt{c_{1-\alpha}^*(\hat{\rho})}$ . In this case, our CI is not empty and is constructed from the average of  $\hat{\theta}_l, \hat{\theta}_u$ . Intuitively, this accounts for the case where  $\hat{\theta}_l$  is larger than  $\hat{\theta}_u$ , because  $\theta_l = \theta_u$ . In this case, it is known that the ‘optimal’ estimator of the common value  $\theta_l$  or  $\theta_u$  is, as is evident in the original definition of  $A$  and  $B$ , a weighted average of the two estimators  $\hat{\theta}_l, \hat{\theta}_u$  and our CI automatically makes use of the ‘optimal’ estimator. Appendix C provides the detailed derivation.

**Remark 1.** (i) It is easy to see that  $CI_{\text{FP}}$  is nested; (ii) It is straightforward to extend  $CI_{\text{FP}}$  with  $c_{1-\alpha}^*(\hat{\rho})$  defined in (7) to the case where the asymptotic distribution of  $(\hat{\theta}_l, \hat{\theta}_u)$  is non-normal, as long as it 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 parameters in the model. Park (2007a) is working on an extension of  $CI_{\text{FP}}$  to inference for the distribution of the treatment effects for randomized data.

**Remark 2.** It follows from the proof of Theorem 2.1 that  $CI_{\text{FP}}$  remains to be asymptotically uniformly valid and non-conservative even when  $\Delta$  is a known but non-zero constant or when  $\Delta$  is bounded away from zero. In contrast,  $CI_{\text{PA}}$  is conservative when  $\Delta$  is a known but non-zero constant or when  $\Delta$  is bounded away from zero.

### 2.3 A Comparison of the New CI with the CI of Stoye (2007)

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

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

Then a CI for  $\theta_0$  can be defined as

$$\begin{aligned} \overline{CI}_S &= \{ \theta : T_{nl}(\theta) \leq c_l \wedge T_{nu}(\theta) \leq c_u \} \\ &= \begin{cases} \left[ \hat{\theta}_l - \frac{\sqrt{c_l} \hat{\sigma}_l}{\sqrt{n}}, \hat{\theta}_u + \frac{\sqrt{c_u} \hat{\sigma}_u}{\sqrt{n}} \right] & \text{if } \hat{\theta}_l - \frac{\sqrt{c_l} \hat{\sigma}_l}{\sqrt{n}} \leq \hat{\theta}_u + \frac{\sqrt{c_u} \hat{\sigma}_u}{\sqrt{n}} \\ \emptyset & \text{otherwise} \end{cases}, \end{aligned} \quad (12)$$

where  $c_l, c_u$  are chosen to guarantee the correct level of coverage.<sup>2</sup> (12) reveals that  $\overline{CI}_S$  is of the same form as the CI proposed by Stoye (2007). Note that under  $\{\gamma_{\omega_n, h}\}$ ,

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

We obtain

$$\begin{aligned} & \inf_{h_l \geq 0, h_u \geq 0, \hat{\sigma}_u h_u + \hat{\sigma}_l h_l = \sqrt{n} \Delta^*} \Pr(\theta \in \overline{CI}_S) \\ &= \Pr(Z_{l, h_\rho} \leq h_l + \sqrt{c_l} \wedge Z_{u, h_\rho} \geq -h_u - \sqrt{c_u}) \\ &= \min \left\{ \begin{array}{l} \Pr\left(Z_{l, h_\rho} \leq \sqrt{c_l} \wedge Z_{u, h_\rho} \geq -\frac{\sqrt{n} \Delta^*}{\hat{\sigma}_u} - \sqrt{c_u}\right), \\ \Pr\left(\frac{\sqrt{n} \Delta^*}{\hat{\sigma}_l} + Z_{l, h_\rho} \leq \sqrt{c_l} \wedge Z_{u, h_\rho} \geq -\sqrt{c_u}\right) \end{array} \right\} \\ &= \min \left\{ \begin{array}{l} \Phi\left(\sqrt{c_u} + \frac{\sqrt{n} \Delta^*}{\hat{\sigma}_u}\right) - \Phi\left(-\sqrt{c_l}, \sqrt{c_u} + \frac{\sqrt{n} \Delta^*}{\hat{\sigma}_u}; h_\rho\right), \\ \Phi\left(\sqrt{c_u}\right) - \Phi\left(-\sqrt{c_l} - \frac{\sqrt{n} \Delta^*}{\hat{\sigma}_l}, \sqrt{c_u}; h_\rho\right) \end{array} \right\} \end{aligned} \quad (13)$$

where

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

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

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

$$\min \left\{ \begin{array}{l} \Phi\left(\sqrt{c_u} + \frac{\sqrt{n} \Delta^*}{\hat{\sigma}_u}\right) - \Phi\left(-\sqrt{c_l}, \sqrt{c_u} + \frac{\sqrt{n} \Delta^*}{\hat{\sigma}_u}; \hat{\rho}\right), \\ \Phi\left(\sqrt{c_u}\right) - \Phi\left(-\sqrt{c_l} - \frac{\sqrt{n} \Delta^*}{\hat{\sigma}_l}, \sqrt{c_u}; \hat{\rho}\right) \end{array} \right\} = 1 - \alpha.$$

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

<sup>2</sup>We changed the definitions of  $c_l$  and  $c_u$  in (4) to be consistent with other parts in the paper. As a result,  $c_l$  and  $c_u$  in (4) are  $\sqrt{c_l}$  and  $\sqrt{c_u}$  in (12). 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  $\theta_0$  that satisfies the moment conditions:

$$\begin{aligned} Em_j(W_i, \theta_0) &\geq 0 \text{ for } j = 1, \dots, p \text{ and} \\ Em_j(W_i, \theta_0) &= 0 \text{ for } j = p + 1, \dots, p + v, \end{aligned} \tag{14}$$

where  $\{m_j(\cdot, \theta) : j = 1, \dots, p + v\}$  are known real-valued moment functions and  $\{W_i : i \geq 1\}$  are observed i.i.d. random vectors<sup>3</sup> with joint distribution  $P$ . The true value  $\theta_0$  is not necessarily point identified, but the moment equalities/inequalities in (14) restrict the set of values of  $\theta_0$ , referred to as the identified set of  $\theta_0$ . In many economic/econometric models, the parameters of interest are defined by a finite number of moment equalities/inequalities in (14). One widely studied example of partially identified models in microeconomic literature is an entry game with stochastic payoff functions, see Bresnahan and Reiss (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 below provides a brief summary of the inequality moment constraints. For a complete description of this problem, see Tamer (2003) or Moon and Schorfheide (2007).

**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 when both firms are profitable as monopolist but not as duopolist,  $E[Y_1(1 - Y_0)|X_1, X_2] = P(Y_1 = 1, Y_0 = 0|X_1, X_2)$  satisfies

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

where

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

---

<sup>3</sup>The i.i.d. assumption is made for ease of exposition. This can be relaxed, see AG (2007).

Similar bounds can be constructed for  $E[Y_1(1 - Y_0)|X_1, X_2] = P(Y_1 = 0, Y_0 = 1|X_1, X_2)$ . Together they imply moment inequality constraints on the model parameters.

Another example of parameters defined by moment equalities/inequalities is that of regression models with interval outcomes in Manski and Tamer (2002).

**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  $\theta$  satisfies

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

These conditional restrictions imply the inequalities

$$E[Y_{Li}Z_i] \leq \theta' E[X_iZ_i] \leq E[Y_{Ui}Z_i], \quad (15)$$

where  $Z_i$  is a vector of positive transformations of  $X_i$ , see CHT. Let  $Z_i$  be of dimension  $q$ . This falls in the moment inequality framework of (14) with  $p = 2q, v = 0$ , see also CHT, AG (2007), and Beresteanu and Molinari (2006).

Additional examples can be found in the references cited in the Introduction. In general, the identified set for  $\theta_0$  defined in (14) 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  $\theta_0$  in (14) and clarify its relation to existing non-resampling based CSs in Rosen (2005), Soares (2006), PPHI, and AG (2007).

Let

$$m(W_i, \theta) = (m_1(W_i, \theta), \dots, m_k(W_i, \theta))',$$

where  $k = p + v$ . We make the same assumptions as in (3.3) of AG (2007) which are restated as Assumption MI in Appendix A. Define  $\gamma_1 = (\gamma_{1,1}, \dots, \gamma_{1,p})' \in R_+^p$  by writing the moment inequalities in (14) as moment equalities:

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

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

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

where  $\bar{m}_{n,j}(\theta) = n^{-1} \sum_{i=1}^n m_j(W_i, \theta)$  and  $\hat{\sigma}_{n,j}^2(\theta)$  is a consistent estimator of  $\sigma_j^2(\theta)$ . Let  $\Omega = \Omega(\theta) = Corr(m(W_i, \theta))$  and  $\bar{m}_n(\theta) = (\bar{m}_{n,1}(\theta), \dots, \bar{m}_{n,k}(\theta))$ .

Let  $\gamma_2 = (\gamma_{2,1}, \gamma_{2,2}) = (\theta, \text{vech}_*(\Omega))$ , where  $\text{vech}_*(\Omega)$  denotes the vector of elements of  $\Omega$  that lie below the main diagonal, and  $\gamma_3$  the remaining parameters in the model. AG (2007) showed that under the local sequence  $\{\gamma_{\omega_n, h}\}$ ,

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

where  $h = (h_1, h_2)$  in which  $h_1 = \lim(\omega_n^{1/2} \gamma_{\omega_n, h, 1})$  and  $h_2 \equiv (h_{2,1}, h_{2,2}) = \lim(\omega_n^{1/2} \gamma_{\omega_n, h, 2})$ ,  $Z_{h_{2,2}} = (Z_{h_{2,2},1}, \dots, Z_{h_{2,2},k})' \sim N(0_k, \Omega_{h_{2,2}})$  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  $\widehat{D}_n(\theta) = \text{Diag}(\widehat{\Sigma}_n(\theta))$  and

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

Let  $J_h$  denote the distribution function of the random variable  $\sum_{j=1}^p [Z_{h_{2,2},j} + h_1]_-^2 + \sum_{j=p+1}^{p+v} [Z_{h_{2,2},j}]^2$ . 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  $\Omega_{h_{2,2}}$ :  $cv_{1-\alpha}(h_1, \Omega_{h_{2,2}})$ .  $\Omega_{h_{2,2}}$  can be consistently estimated whereas  $h_1$  cannot. To circumvent this problem, AG (2007) proposed a PA-CS for  $\theta_0$  by using the critical value  $cv_{1-\alpha}(0, \widehat{\Omega}_n(\theta))$ . 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  $\gamma_1$  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  $\theta_0$  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 is not asymptotically conservative even when  $\Delta$  is bounded away from zero. However, it does not allow for a straightforward generalization to parameters defined by general moment equalities/inequalities, as there is no such simple characterization of restrictions of this type. Instead we propose the following remedy: for  $j = 1, \dots, p$ , we define

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

Let  $\gamma_1^*(\theta) = (\gamma_{1,1}^*(\theta), \dots, \gamma_{1,p}^*(\theta))$  and define<sup>4</sup>

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

**THEOREM 3.1** *Under the same assumptions as those in Theorem 2 (a) of AG (2007), i.e., Assumption MI stated in Appendix A, we have*

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

**Remark 2.** Like  $CI_{FP}$ ,  $CS_{MC}$  remains to be asymptotically uniformly valid and non-conservative even when the vector of slackness parameters  $\gamma_1$  is bounded away from zero, implying there are restrictions on the moment inequalities such that satisfaction of one inequality as an equality implies violation of another.

It is interesting to observe that the CSs of Rosen (2005), Soares (2006), and the PA-CS of AG (2007) and CHT are all<sup>5</sup> based on  $cv_{1-\alpha} \left( h_1, \widehat{\Omega}_n(\theta) \right)$  except that they use different values of  $h_1$ : PA-CS uses  $cv_{1-\alpha} \left( 0, \widehat{\Omega}_n(\theta) \right)$  and is thus asymptotically conservative when  $\gamma_1$  is bounded away from zero; Rosen (2005) and Soares (2006) use  $cv_{1-\alpha} \left( 0, \dots, 0, \infty, \dots, \infty, \widehat{\Omega}_n(\theta) \right)$  with  $p^*$  zeros, where  $p^*$  is an upper bound on the number of binding inequality constraints in Rosen (2006) and 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 that of Rosen (2005) and the PA-CS. However, as Soares (2006) pointed out, this procedure may be computationally intensive depending on the dimension of  $\theta$ .

**Interval-Identified Parameters.** Instead of estimating  $\Delta = \theta_u - \theta_l$  by the shrinkage estimator  $\Delta^*$ , we can also ‘estimate’  $\gamma_{1l}$  and  $\gamma_{1u}$  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  $\theta_0$  can be defined as follows:

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

Note that the use of shrinkage ‘estimators’  $\gamma_{1l}^*$  and  $\gamma_{1u}^*$  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,

<sup>4</sup>Independently, Andrews and Soares (2007) proposed similar confidence sets in this context. Instead of using  $\sqrt{n}\gamma_1^*(\theta)$  to replace  $h_1$  in  $cv(h_1, \Omega_{h_{2,2}})$ , they used functions of

$$\kappa_n^{-1} \sqrt{n} (\overline{m}_{n,1}(\theta) / \widehat{\sigma}_{n,1}(\theta), \dots, \overline{m}_{n,p}(\theta) / \widehat{\sigma}_{n,p}(\theta)),$$

where  $\kappa_n \rightarrow \infty$  and  $\kappa_n^{-1} \sqrt{n} \rightarrow \infty$  as  $n \rightarrow \infty$ .

<sup>5</sup>Rosen (2005) uses a different test statistic from  $T_n(\theta)$ .

$\gamma_{1l}^* = 0$ . In contrast,  $\gamma_{1u}^*$  would approach  $\Delta/\sigma_u$ . At the boundaries, the two CSs:  $CI_{FP}$  and  $CS_{IP}$  behave similarly.

**Regression Models with Interval Outcomes.** Obviously,  $CS_{MC}$  is valid for regression models with interval outcomes. In addition, if  $q = 1$ , we can also extend  $CI_{FP}$  to  $\theta_0$ . Let  $W_i = (Y_{Li}, Y_{Ui}, X_i, Z_i)$ ,

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

Let

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

where  $\rho(\theta) = \text{Corr.}(m_1(W_i, \theta), m_2(W_i, \theta))$ . Let  $J_{(h_l, h_u, \rho)}$  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

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

Let  $cv_{1-\alpha}(h_l, h_u, \rho)$  denote the  $1-\alpha$  quantile of  $J_{(h_l, h_u, \rho)}$ . An alternative CS for  $\theta_0$  uses the following critical value:

$$c_{1-\alpha}(\theta) = cv_{1-\alpha} \left( \frac{\sqrt{n}\Delta^*}{\max\{\widehat{\sigma}_{n,1}(\theta), \widehat{\sigma}_{n,2}(\theta)\}}, 0, \widehat{\rho}(\theta) \right), \quad (16)$$

where  $\Delta^*$  is a shrinkage estimator of  $\Delta$  defined as

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

and

$$\widehat{\rho}(\theta) = \frac{n^{-1} \sum_{i=1}^n [m_1(W_i, \theta) - \overline{m}_{n,1}(\theta)] [m_2(W_i, \theta) - \overline{m}_{n,2}(\theta)]}{\widehat{\sigma}_{n,1}(\theta) \widehat{\sigma}_{n,2}(\theta)}.$$

## 4 Numerical Studies

In this section, we first present a numerical comparison of the critical values of four CIs at 0.95 nominal level:  $CI_{FP}$ ,  $CI_S$ ,  $CI_{PA}$ , 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  $CI_{PA}$ .

### 4.1 Comparison of Critical Values

The CIs:  $CI_{PA}$  and  $CI_{IM}$  are respectively based on  $cv_{1-\alpha}(0, 0, \rho)$  and  $\sqrt{cv_{1-\alpha}(0, 0, 1)}$ . Let  $\alpha = 0.05$ . 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  $\rho$  increases and approaches to  $\Phi^{-1}(1 - \alpha/2) = 1.96$  as  $\rho \rightarrow 1$ . But for small values of  $\rho$ ,  $cv_{1-\alpha}(0, 0, \rho)$  can be much larger than  $cv_{1-\alpha}(0, 0, 1)$ . If  $\Delta$  is bounded away from zero, it follows from the proof of Theorem 2.1 and the monotonicity of  $\sqrt{cv_{0.95}(0, 0, \rho)}$  that

$$\lim_{n \rightarrow \infty} \inf_{\theta \in \Theta} \inf_{P: \theta_0(P) = \theta} \Pr(\theta_0 \in CS_{PA}) = \Pr(Z_{l, \rho} \leq \sqrt{cv_{0.95}(0, 0, 1)}) = 0.975.$$

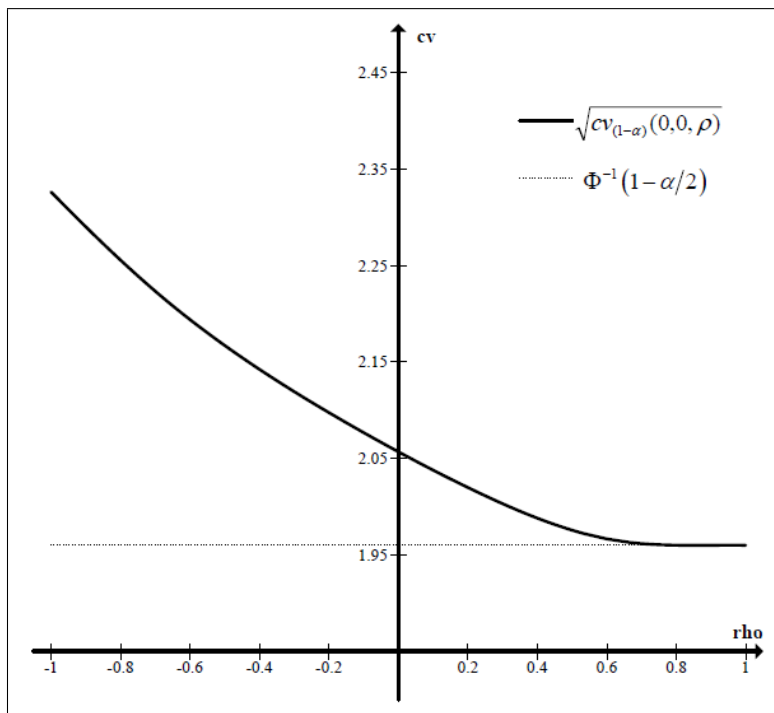


Figure 1.  $\sqrt{cv_{0.95}(0, 0, \rho)}$  and  $\Phi^{-1}(0.975)$

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

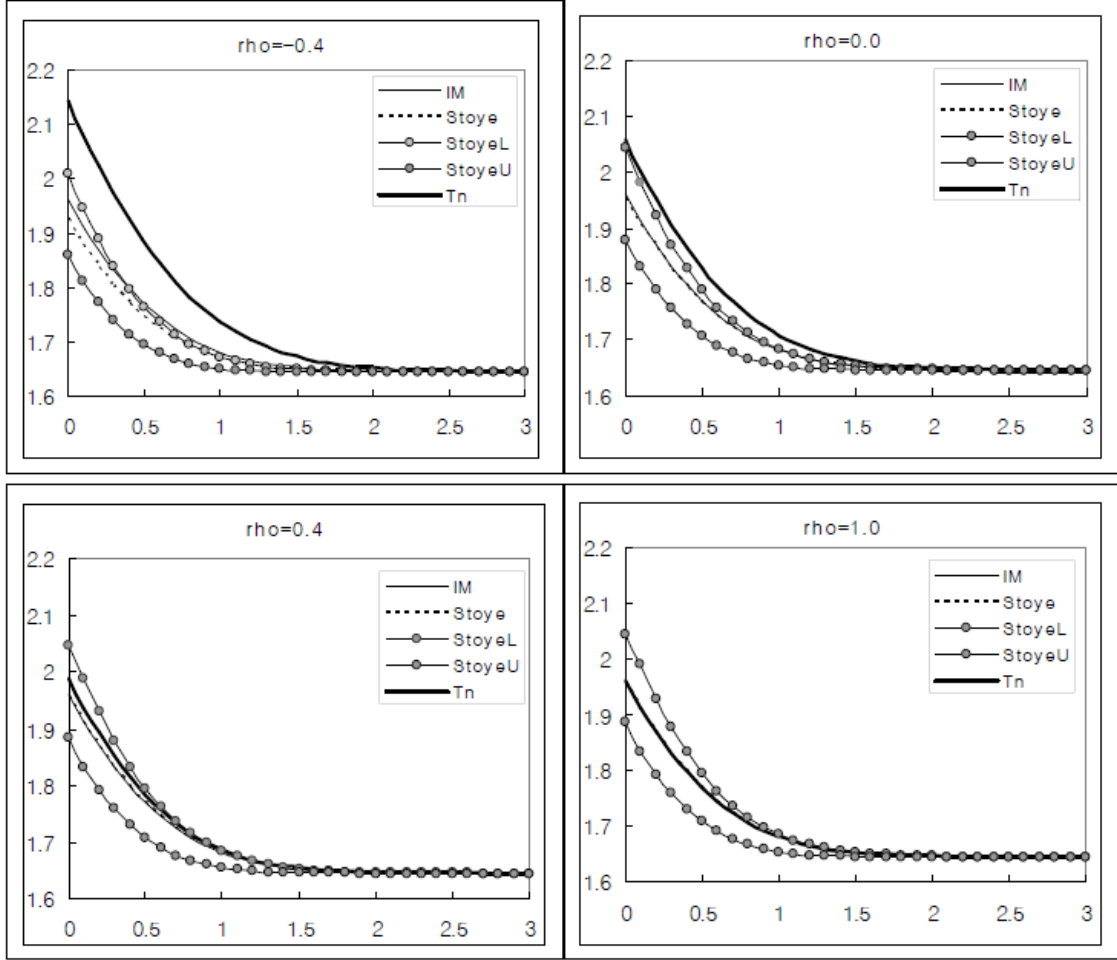


Figure 2. Comparison of critical values

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

Several interesting conclusions can be made based on Figure 2. First, when  $\sqrt{n}\Delta / \max\{\sigma_l, \sigma_u\} > 2.5$ , all the critical values become almost identical to  $\Phi^{-1}(1 - \alpha) = 1.645$ . Second, when  $\sqrt{n}\Delta / \max\{\sigma_l, \sigma_u\}$  is small, the critical values for different CIs differ and the difference becomes larger as  $\rho$  approaches  $-1$ . Third, when  $\rho$  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  $CI_{PA}$  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
$\exp(Y)$ ( <i>wages and salaries</i> , 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  $\hat{\theta}_u = n^{-1} \sum_i Y_{Ui}$ .

We considered three DGPs designed to shed light on the performance of  $CI_{FP}$ ,  $CI_S$ , and  $CI_{PA}$  in three typical cases: point-identified case, interval identified case with a small  $\Delta$ , and interval identified case with a large  $\Delta$ . 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  $i = 1, \dots, n$  and the other with  $\{Y_{Li}\}_{i=1}^n, \{Y_{Ui}\}_{i=1}^n$  being independent. For interval identified case with small  $\Delta$ , 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  $\Delta$ , 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	$[\theta_l, \theta_u]$	$[\sigma_l, \sigma_u]$	$\rho$	$\Delta$
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, 7.5234]	[1.830, 1.440]	1.0	3.7251

The length of the identified interval  $\Delta$  in the 16 bracket case is eight times smaller than that of the 2-bracket case. Moreover, the magnitude of  $\Delta$  in the 16 bracket experiment is almost half of  $\sigma_l$  and  $\sigma_u$ . So,  $\theta_l$  and  $\theta_u$  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,  $\Delta$  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  $b_n = s.d.(\hat{\Delta})c/\ln(n)$  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  $\theta_0$  is point identified or when  $\Delta$  is small, it's possible that  $\hat{\theta}_l$  is larger than  $\hat{\theta}_u$  in which case, the effect of using the shrinkage estimator with  $b_n = 0$  is to replace negative  $\hat{\Delta}$ 's with zero; Second, when  $\Delta$  is large enough, the shrinkage estimator with  $b_n = 0$  is the same as the original estimator and in this case, we'll observe the performance of  $CI_{FP}$  and  $CI_S$  using the original estimator  $\hat{\Delta}$ . When  $c = 3.5, 4$ ,  $b_n$  satisfies the conditions of Theorem 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  $CI_{PA}$  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  $\theta_l$  ( $h_l = 0$ ) or  $\theta_u$  ( $h_u = 0$ ), we report the respective coverage rates of  $CI_{FP}$ ,  $CI_S$ , and  $CI_{PA}$  for  $\theta = \theta_l, \theta_u$ .

#### 4.2.1 Point-Identified Case

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

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

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

Table 3: Summary Statistics for  $CI_n$

$n$	$CR(\theta_0)$	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,  $\hat{\Delta}$  may not be exactly zero. In fact, it is possible that  $\hat{\Delta}$  is negative. Since we drew random samples  $\{Y_{Li}\}$  and  $\{Y_{Ui}\}$  independently, we would expect this to happen at about 50% of the simulations. In Table 4 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  $\Delta^*$  plays a role. When  $c = 0$ ,  $P(\Delta^*)$  shows the proportion of simulations with negative  $\hat{\Delta}$ . It is about 0.5 for all sample sizes. In addition, we reported the coverage rates and width of each CI based on each value of  $b_n$  together with the average of  $\sqrt{c_{1-\alpha}}$  denoted as  $\text{Avg}(\sqrt{c_{1-\alpha}})$ <sup>6</sup>.

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

$n$		$c$	$P(\Delta^*)$	$\text{Avg}(\sqrt{c_{1-\alpha}})$	$\text{CR}(\theta_0)$	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
	$CI_{PA}$			2.0569	0.9480	0.1833
	1000	$CI_S$	0	0.4945	(1.8476, 1.8318)	0.9425
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
$CI_{PA}$				2.0569	0.9445	0.1298
2000		$CI_S$	0	0.496	(1.8459, 1.8323)	0.9455
	(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
	$CI_{PA}$			2.0569	0.9425	0.0915
	8000	$CI_S$	0	0.499	(1.844, 1.833)	0.9470
(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
$CI_{PA}$				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  $CI_{PA}$ , 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.

<sup>6</sup>For  $CI_S$ , we provide  $(\sqrt{c_{l,1-\alpha}}, \sqrt{c_{u,1-\alpha}})$  which correspond to  $(c_{l,1-\alpha}, c_{u,1-\alpha})$  in the original Stoye's notation.

## 4.2.2 Interval-Identified Case

**Sixteen Brackets: A small  $\Delta$**  The coverage rates for  $\theta_l$  and  $\theta_u$  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( $\theta_l$ )	CR( $\theta_u$ )
500	$CI_S$	0	0	(1.6449, 1.6449)	0.6082	<b>0.9235</b>	0.9360
		(3.5, 4)	1	(1.9024, 2.0263)	0.6353	<b>0.9550</b>	0.9725
	$CI_{FP}$	0	0	1.6449	0.6082	<b>0.9235</b>	0.9360
		(3.5, 4)	1	1.9759	0.6371	<b>0.9595</b>	0.9655
	$CI_{PA}$			1.9759	0.6371	<b>0.9595</b>	0.9655
	1000	$CI_S$	0	0	(1.6449, 1.6449)	0.5653	<b>0.9230</b>
3.5, 4			1	(1.9020, 2.0260)	0.5845	<b>0.9535</b>	0.9715
$CI_{FP}$		0	0	1.6449	0.5653	<b>0.9230</b>	0.9340
		(3.5, 4)	1	1.9760	0.5857	<b>0.9570</b>	0.9630
$CI_{PA}$				1.9760	0.5857	<b>0.9570</b>	0.9630
2000		$CI_S$	0	0	(1.6449, 1.6449)	0.5367	<b>0.9335</b>
	3.5		0.4655	(1.7641, 1.8228)	0.5429	<b>0.9515</b>	0.9625
	4		1	(1.9015, 2.0263)	0.5503	<b>0.9570</b>	0.9685
	$CI_{FP}$	0	0	1.6449	0.5367	<b>0.9335</b>	0.9370
		3.5	0.4655	1.7990	0.5433	<b>0.9570</b>	0.9580
		4	1	1.9761	0.5512	0.9640	<b>0.9630</b>
$CI_{PA}$			1.9761	0.5512	0.9640	<b>0.9630</b>	
8000	$CI_S$	(0, 3.5, 4)	0	(1.6449, 1.6449)	0.5013	0.9450	<b>0.9435</b>
	$CI_{FP}$	(0, 3.5, 4)	0	1.6449	0.5013	0.9450	<b>0.9435</b>
	$CI_{PA}$			1.9761	0.5086	0.9720	<b>0.9705</b>

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  $\Delta^*$ . 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  $CI_{PA}$ . 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  $CI_{PA}$ , 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  $\Delta$**  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( $\theta_l$ )	CR( $\theta_u$ )
500	$CI_S$	(1.6449, 1.6449)	3.9655	<b>0.9435</b>	0.9580
	$CI_{FP}$	1.6449	3.9655	<b>0.9435</b>	0.9580
	$CI_{PA}$	1.960	4.0115	<b>0.9655</b>	0.9775
1000	$CI_S$	(1.6449, 1.6449)	3.8949	<b>0.9455</b>	0.9495
	$CI_{FP}$	1.6449	3.8949	<b>0.9455</b>	0.9495
	$CI_{PA}$	1.960	3.8949	<b>0.9685</b>	0.9785
2000	$CI_S$	(1.6449, 1.6449)	3.8453	<b>0.9480</b>	0.9495
	$CI_{FP}$	1.6449	3.8453	<b>0.9480</b>	0.9495
	$CI_{PA}$	1.960	3.8453	<b>0.9680</b>	0.9745
8000	$CI_S$	(1.6449, 1.6449)	3.8753	<b>0.9465</b>	0.9515
	$CI_{FP}$	1.6449	3.8753	<b>0.9465</b>	0.9515
	$CI_{PA}$	1.960	3.8753	0.9760	<b>0.9735</b>

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,  $CI_{PA}$  is quite different from  $CI_S$  and  $CI_{FP}$ : it overcovers for all sample sizes. Secondly, the critical value for  $CI_{PA}$  is  $\Phi^{-1}(1 - \alpha/2) = 1.96$ , because  $\hat{\rho} = 1$ ; while that for  $CI_S$  and  $CI_{FP}$  is  $\Phi^{-1}(1 - \alpha) = 1.645$ , 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 two-sided 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  $(\widehat{\theta}_l, \widehat{\theta}_u)$  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 applications.

Park (2007a) investigates inference for the distribution of the treatment effects of a binary treatment. Using the same notation as in Example 2, but define  $\theta_0 = F_{\Delta}(\delta)$ ,  $\theta_l = \sup_y \max(F_1(y) - F_0(y - \delta), 0)$  and  $\theta_u = 1 + \inf_y \min(F_1(y) - F_0(y - \delta), 0)$ . Then it is known that  $\theta_l \leq \theta_0 \leq \theta_u$ . Again, with randomized data,  $F_1$  and  $F_0$  are identified and thus  $\theta_l$ ,  $\theta_u$  are identified. Estimators of  $\theta_l, \theta_u$  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  $\theta_l, \theta_u$  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  $\inf_y(F_1(y) - F_0(y - \delta))$ . Fan and Park (2007b) considered inference on the bounds themselves.

Another example violating Assumption IM (i) concerns the ‘mixing problem’ discussed by Manski (1997, 2003). The ‘mixing problem’ arises, for example, when we want to “extrapolate the results from a randomized experiment,” see Manski (2003). Since we do not know the ‘treatment shares,’ i.e., the possibility that people comply the rule and do not, the probability for a certain range of outcomes, say  $y \in B$ , to occur lies in  $[\max\{P_1(y \in B) + P_0(y \in B) - 1, 0\}, \min\{P_1(y \in B) + P_0(y \in B), 1\}]$ , where  $P_j$ ,  $j = 1, 0$ , is the probability measure corresponding to  $F_j$ . Park (2007c) studies the statistical inference for this problem and provides some empirical applications.

Park (2007b) provides an application of the tools developed in Fan and Park (2007b) and Park (2007a, 2007c) 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 documented in the literature (see e.g., Ding and Lehrer 2005), it has not been fully investigated empirically.

## 6 Appendix A: Technical Proofs

For convenience, we restate the assumptions (3.3) in AG (2007) as Assumption MI below.

**Assumption MI.** For i.i.d. observations, the parameter space for  $(\theta, P)$  is the set of all  $(\theta, P)$  that satisfy:

- (i)  $Em_j(W_i, \theta_0) \geq 0$  for  $j = 1, \dots, p$ ,
- (ii)  $Em_j(W_i, \theta_0) = 0$  for  $j = p + 1, \dots, k$ ,
- (iii)  $\{W_i\}_{i=1}^n$  are i.i.d.,
- (iv)  $\sigma_j^2(\theta) \in (0, \infty)$  for  $j = 1, \dots, k$ ,
- (v)  $Corr(m(W_i, \theta)) \in \Psi$ , and
- (vi)  $E|m_j(W_i, \theta) / \sigma_j(\theta)|^{2+\delta} \leq M$  for  $j = 1, \dots, k$ ,

where  $\Psi$  is the set of correlation matrices, and  $M < \infty, \delta > 0$  are fixed constants.

**Proof of Theorem 2.1.** 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 with  $J_h$  the distribution function of the random variable  $(Z_{l, h_\rho} - h_l)_+^2 + (Z_{u, h_\rho} + h_u)_-^2$ . Similar to Stoye (2007), we let  $c_n = (n^{-1/2}b_n)^{1/2}$ . Then  $c_n \rightarrow 0$  and  $n^{1/2}c_n \rightarrow \infty$ . We consider two cases: Case I.  $\Delta_n \geq c_n$ ; Case II.  $\Delta_n < c_n$ .

**Case I.**  $\Delta_n \geq c_n$ . In this case,  $n^{1/2}\Delta_n \geq n^{1/2}c_n \rightarrow \infty$ , so either  $h_l = \infty$  or  $h_u = \infty$  or both. Suppose  $h_l = \infty$ . Then under the local sequence  $\{\gamma_{\omega_n, h}\}$ , we obtain

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

where we have used the result that the random variable  $(Z_{u, h_\rho} + h_u)_-^2$  is stochastically decreasing in  $h_u \geq 0$  and the result that  $\Pr[\Delta^* = \hat{\Delta}] \rightarrow 1$  because  $\Pr[\hat{\Delta} > b_n] \rightarrow 1$ . The proof for  $h_u = \infty$  is similar. Suppose both  $h_l = \infty$  and  $h_u = \infty$ . Then it is easy to see that  $\Pr[\theta \in CI_{\text{FP}}] \rightarrow 1$ .

**Case II.**  $\Delta_n < c_n$ . In this case, Stoye (2007) shows that  $\Delta^* = 0 \leq \Delta$  with probability approaching one. Note that under the local sequence  $\{\gamma_{\omega_n, h}\}$ ,

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

where we have used the result that the random variable  $(Z_{l, h_\rho} - h_l)_+^2 + (Z_{u, h_\rho} + h_u)_-^2$  is stochastically decreasing in  $h_l \geq 0, h_u \geq 0$ . The proof is completed by noting that when  $\Delta = 0$ ,  $\Pr[\theta \in CI_{\text{FP}}] \rightarrow 1 - \alpha$ .

**Proof of Theorem 3.1.** We prove the result when  $p = 2$ . The general case is similar. Similar to the proof of Theorem 2.1, we need to justify the use of  $\gamma_1^*(\theta) = (\gamma_{1,1}^*(\theta), \gamma_{1,2}^*(\theta))$ , where

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

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

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

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

**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,

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

**Case III.** 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  $n^{1/2}\gamma_{1,2}(\theta) \geq n^{1/2}c_n \rightarrow \infty$ . Thus,

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

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

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

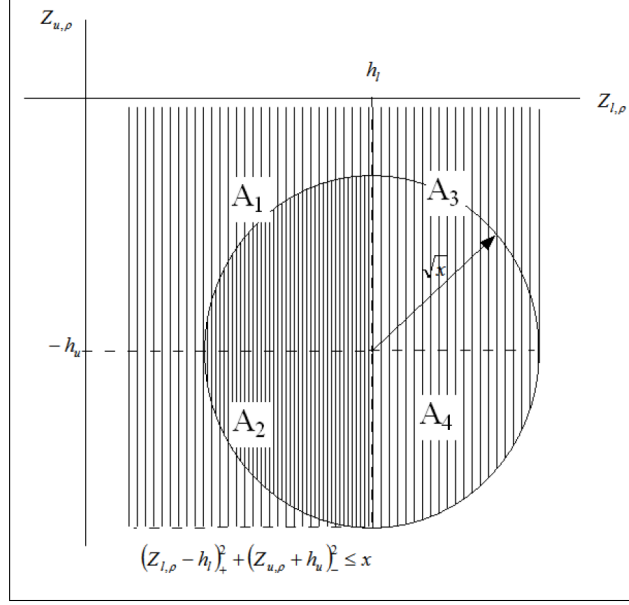
In this section, we derive a closed-form expression for  $J_h(x)$ . This should be useful in constructing CSs in moment inequality models when there are two moment constraints. Let  $\phi(z_l, z_u; \rho)$  and  $\Phi(z_l, z_u; \rho)$  denote respectively the pdf and cdf of  $(Z_{l,\rho}, Z_{u,\rho})$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

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

If  $|\rho| < 1$ , then

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

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



Hence,

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

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

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

Therefore, we get

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

If  $\rho = -1$ , then

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

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

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

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

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

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

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

Also,

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

Therefore, we get

$$B_1(x) = \begin{cases} \left\{ Z : h_{\max} < Z \leq \frac{h_{\max} + h_{\min} + \sqrt{2x - (h_{\max} - h_{\min})^2}}{2} \right\} & \text{if } x > (h_{\max} - h_{\min})^2, \\ \emptyset & \text{otherwise} \end{cases}$$

Similarly, we can show:

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

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

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

Combining them altogether, we get

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

Therefore,

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

## 8 Appendix C. The Forms of $CI_{PA}$ and $CI_{FP}$

In this section, we show that both  $CI_{PA}$  and  $CI_{FP}$  are intervals because their critical values do not depend on  $\theta$ . In general,  $CS_n$  defined as

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

with a constant critical value  $c_{1-\alpha}$  has the following alternative expressions:

$$CS_n = \begin{cases} \left[ \hat{\theta}_l - \sqrt{c_{1-\alpha}} \frac{\hat{\sigma}_l}{\sqrt{n}}, \hat{\theta}_u + \sqrt{c_{1-\alpha}} \frac{\hat{\sigma}_u}{\sqrt{n}} \right] & \text{if } \sqrt{n} \hat{\Delta} \geq -\sqrt{c_{1-\alpha}} \min \{ \hat{\sigma}_l, \hat{\sigma}_u \} \\ \left[ \hat{\theta}_l - \sqrt{c_{1-\alpha}} \frac{\hat{\sigma}_l}{\sqrt{n}}, B \right] & \text{if } -\sqrt{c_{1-\alpha}} \hat{\sigma}_l \leq \sqrt{n} \hat{\Delta} < \sqrt{c_{1-\alpha}} \hat{\sigma}_u \\ \left[ A, \hat{\theta}_u + \sqrt{c_{1-\alpha}} \frac{\hat{\sigma}_u}{\sqrt{n}} \right] & \text{if } -\sqrt{c_{1-\alpha}} \hat{\sigma}_u \leq \sqrt{n} \hat{\Delta} < \sqrt{c_{1-\alpha}} \hat{\sigma}_l \\ [A, B] & \text{if } -\sqrt{c_{1-\alpha}} (\hat{\sigma}_u^2 + \hat{\sigma}_l^2) \leq \sqrt{n} \hat{\Delta} < -\sqrt{c_{1-\alpha}} \max \{ \hat{\sigma}_l, \hat{\sigma}_u \} \\ \emptyset & \text{if } \sqrt{n} \hat{\Delta} < -\sqrt{c_{1-\alpha}} (\hat{\sigma}_l^2 + \hat{\sigma}_u^2) \end{cases} \quad (17)$$

where

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

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

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

For **Case II**, one may think

$$CS_n = \left( -\infty, \hat{\theta}_u + \sqrt{c_{1-\alpha}} \frac{\hat{\sigma}_u}{\sqrt{n}} \right] \cap \left[ \hat{\theta}_l - \sqrt{c_{1-\alpha}} \frac{\hat{\sigma}_l}{\sqrt{n}}, \infty \right).$$

However, it is more complicated. We'll examine it in detail. Note that

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

where

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

By definition, we obtain

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

and

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

Now,

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

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

$$CS_{n3} = \begin{cases} [\max \{A, \hat{\theta}_u\}, \min \{B, \hat{\theta}_l\}] & \text{if } n\hat{\Delta}^2 \leq c_{1-\alpha} (\hat{\sigma}_l^2 + \hat{\sigma}_u^2) \\ \emptyset & \text{otherwise} \end{cases}.$$

Now, one can show:

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

and

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

Summarizing, we get

$$\begin{aligned}
CS_n &= CS_{n1} \cup CS_{n3} \cup CS_{n2} \\
&= \begin{cases} \left[ \hat{\theta}_l - \sqrt{c_{1-\alpha} \frac{\hat{\sigma}_l}{\sqrt{n}}}, \hat{\theta}_u + \sqrt{c_{1-\alpha} \frac{\hat{\sigma}_u}{\sqrt{n}}} \right] & \text{if } n\hat{\Delta}^2 \leq c_{1-\alpha} \min \{ \hat{\sigma}_l^2, \hat{\sigma}_u^2 \} \\ \left[ \hat{\theta}_l - \sqrt{c_{1-\alpha} \frac{\hat{\sigma}_l}{\sqrt{n}}}, B \right] & \text{if } c_{1-\alpha} \hat{\sigma}_u^2 < n\hat{\Delta}^2 \leq c_{1-\alpha} \hat{\sigma}_l^2 \\ \left[ A, \hat{\theta}_u + \sqrt{c_{1-\alpha} \frac{\hat{\sigma}_u}{\sqrt{n}}} \right] & \text{if } c_{1-\alpha} \hat{\sigma}_l^2 < n\hat{\Delta}^2 \leq c_{1-\alpha} \hat{\sigma}_u^2 \\ [A, B] & \text{if } c_{1-\alpha} \max \{ \hat{\sigma}_l^2, \hat{\sigma}_u^2 \} < n\hat{\Delta}^2 \leq c_{1-\alpha} (\hat{\sigma}_u^2 + \hat{\sigma}_l^2) \\ \emptyset & \text{if } n\hat{\Delta}^2 > c_{1-\alpha} (\hat{\sigma}_l^2 + \hat{\sigma}_u^2) \end{cases}.
\end{aligned}$$

## References

- [1] Andrews, D.W.K. and P. Guggenberger (2005a), “The Limit of Finite-Sample Size and a Problem with Subsampling,” 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] Andrews, D.W.K. and G. Soares (2007), “Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection,” Manuscript, Yale University.
- [6] Beresteanu, A. and F. Molinari (2006), “Asymptotic Properties for a Class of Partially Identified Models,” Manuscript.
- [7] Berry, S.T. (1992), “Estimation of a Model of Entry in the Airline Industry”, *Econometrica* 60, 889-917.
- [8] Bresnahan, T. and P. Reiss (1991), “Empirical models of discrete games,” *Journal of Econometrics* 48, 57–82.
- [9] Bugni, F. A. (2007), “Bootstrap Inference in Partially Identified Models,” mimeo, Northwestern University.
- [10] Canay, I.A. (2007), “EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity,” Manuscript, University of Wisconsin.
- [11] Chernozhukov, V., H. Hong, and E. Tamer (2004), “Inference on Parameter Sets in Econometric Models,” Working Paper, MIT.
- [12] Chernozhukov, V., H. Hong, and E. Tamer (2007), “Parameter Set Inference in a Class of Econometric Models,” *Econometrica* 75, 1243-1284.
- [13] Ciliberto, F. and E. Tamer (2004), “Market Structure and Multiple Equilibria in Airline Markets,” mimeo, Northwestern University.

- [14] 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.
- [15] Fan, Y. and S. Park (2007a), “Statistical Inference on the Quantile of the Treatment effects,” in progress.
- [16] Fan, Y. and S. Park (2007b), “Sharp Bounds on the Distribution of the Treatment effects and Their Statistical Inference,” Unpublished Manuscript, Vanderbilt University.
- [17] Fan, Y. and J. Wu (2007), “Sharp Bounds on the Distribution of the Treatment effects in Switching Regimes Models,” Unpublished Manuscript, Vanderbilt University.
- [18] Galichon, A. and M. Henry (2006), “Inference in Incomplete Models,” Manuscript, Columbia University.
- [19] 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.
- [20] Imbens, G. W. and C. F. Manski (2004), “Confidence Intervals For Partially Identified Parameters,” *Econometrica* 72, 1845–1857.
- [21] Manski, C.F. (1997), “The Mixing Problem in Program Evaluation,” *Review of Economic Studies* 64, 537-553.
- [22] Manski, C. F. (2003), *Partial Identification of Probability Distributions*. Springer.
- [23] Manski, C.F. and E. Tamer (2002), “Inference on Regressions with Interval Data on a Regressor or Outcome,” *Econometrica* 70, 519-546.
- [24] Moon, R. and F. Schorfheide (2007), “A Bayesian Look at Partially-Identified Models,” Manuscript, U.Penn.
- [25] Pakes, A., J. Porter, K. Ho, and J. Ishii (2006), “Moment Inequalities and Their Application,” Working Paper, Harvard University.
- [26] Park, S. (2007a), “Confidence Intervals for the Distribution of the Treatment effects,” in progress.
- [27] Park, S. (2007b), “Heterogeneous Effects of Class Size Reduction: Re-visit the Project STAR,” in progress.
- [28] Park, S. (2007c), “The Mixing Problem in Program Evaluation: Inference,” in progress.

- [29] 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.
- [30] 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.
- [31] Rosen, A.M. (2005), “Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities,” mimeo, University College London.
- [32] Soares, G. (2006), “Inference for Partially Identified Models with Inequality Moment Constraints,” Working Paper, Yale.
- [33] Stoye, J. (2007), “More on Confidence Intervals for Partially Identified Parameters,” Working paper, NYU.
- [34] Vijverberg, W. P. M. (1993), “Measuring the Unidentified Parameter of the Extended Roy Model of Selectivity,” *Journal of Econometrics* 57, 69-89.
- [35] Tamer, E. (2003), “Incomplete Simultaneous Discrete Response Model with Multiple Equilibria,” *Review of Economic Studies*, 70, 147-167.
- [36] 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.