Economics Bulletin

Volume 31, Issue 3

Computationally efficient approximation for the double bootstrap mean bias correction

Rachida Ouysse
The University of New South Wales

Abstract

We propose a computationally efficient approximation for the double bootstrap bias adjustment factor without using the inner bootstrap loop. The approximation converges in probability to the population bias correction factor. We study the finite sample properties of the approximation in the context of a linear instrumental variable model. In identified versions of the model considered in our Monte Carlo experiments, the proposed approximation leads to estimators with lower variance than those based on the double bootstrap and, lower adjusted mean-squared error than estimators based on the single bootstrap. Evidence from the experiments we consider suggests that the bootstrap is less effective in reducing the bias when the instrumental variable is weak and endogeneity is strong.

The author is very grateful to the associate editor and anonymous referee for their helpful comments which helped improve the quality of the paper.
Citation: Rachida Ouysse, (2011) "Computationally efficient approximation for the double bootstrap mean bias correction", *Economics Bulletin*, Vol. 31 no.3 pp. 2388-2403.
Submitted: Apr 28 2010.   **Published:** August 26, 2011.

Computationally efficient approximation for the double bootstrap mean bias correction

Rachida Ouysse*
University of New South Wales
Sydney, Australia

Abstract

We propose a computationally efficient approximation for the double bootstrap bias adjustment factor without using the inner bootstrap loop. The approximation converges in probability to the population bias correction factor. We study the finite sample properties of the approximation in the context of a linear instrumental variable model. In identified versions of the model considered in our Monte Carlo experiments, the proposed approximation leads to estimators with lower variance than those based on the double bootstrap and, lower adjusted mean-squared error than estimators based on the single bootstrap. Evidence from the experiments we consider suggests that the bootstrap is less effective in reducing the bias when the instrumental variable is weak and endogeneity is strong.

KEYWORDS: Bias function; population and sample moments; true data generating process; empirical distribution; Monte Carlo simulation

---

*email: `rouysse@unsw.edu.au`.

1 Introduction

Many econometric estimators although consistent in large samples have small sample bias. The bootstrap is a simulation-based alternative to asymptotic approximation used to correct for finite sample bias for the purpose of point estimation and/or inference.

Advances in statistical theory show that iterating (pre-pivoting) the bootstrap principle brings yet further improvements upon the single bootstrap. Beran (1988) argues that pre-pivoting reduces the dependence between the probability distribution of the resample and the unknown data generating process. Therefore, resampling reinforces the conditions under which the bootstrap performs the best: pivotal or asymptotically pivotal statistics. As a result, the double bootstrap has typically higher order accuracy than the ordinary single bootstrap and the bootstrap can be iterated to reduce the bias by a factor of $O(n^{-1})$ successively. See among others Beran (1988, 1987, 1990), Hall (1992, 1986), Hall and Martin (1988), Lee and Young (1999) and Shi (1992).

These refinements come with a heavy computational cost due to the increasing computational intensity of compounded sampling. This has prompted a number of authors to develop computationally efficient and cheaper alternatives to eliminate the need for nested levels of resampling. Much of the literature is however concerned with ways to generate fast approximations to the P value and quantile functions. To the best of our knowledge, this paper is perhaps the first to address the computational efficiency of approximating the bias function.

The technique of this paper adapts the *fast double* bootstrap of Davidson and MacKinnon (2002b, 2007) and the *warp-speed* method of Giacomini et al. (2007) for approximating the rejection and coverage probabilities of bootstrap tests and confidence intervals to the problem of approximating the bootstrap bias. The approximation requires only twice as many computations as what is usually needed to perform the single bootstrap. We show an optimality result which holds under general conditions and does not require an asymptotic pivot.

The statistical properties of the proposed fast approximation are examined in a linear instrumental variable framework through a Monte Carlo analysis. The results in our experiments show that the fast approximation achieves significant bias reduction over the single bootstrap without the increased variance and the computational cost of the double bootstrap.

We use the following notation throughout the paper: E_μ is the mathematical expectation under the data generating process (DGP) μ, V_μ is the variance under the DGP μ, the indicator function $\mathbf{1}\{x\}$ takes the value 1 if the statement in its argument is correct and 0 otherwise.

2 Bootstrap methods for bias correction

Let $X_1, X_2, ...$ be a sequence of stationary random variables generated from a data generating process μ_0 with unknown joint probability distribution F_0 and possible indexed by an unknown real-valued parameter θ_{μ_0}. Consider a random sample from the data generating process μ_0 of $\mathbf{X}$ with realization $\mathbf{X}^n = (X_1, X_2, ..., X_n)$. Let $\widehat{\mu}$ be the data generating process (DGP) governed by some estimate $\widehat{F}$ of the empirical distribution implied by $\mathbf{X}^n$. We choose the standard uniform $\widehat{F}(y) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{X_i \leq y\}$.

Many statistical problems can be formulated as specifying the statistical properties of the random variable $R_n\left(\mathbf{X}^n, \theta(\mu_0)\right)$ such as its probability distribution function, moments and quantile functions. The bootstrap uses a nonparametric estimate F_n of F_0 to approximate the distribution of R_n using $R_n^* = R_n(\mathbf{X}^n, F_n)$.

The bootstrap principle approximates the sampling distribution of $R_n\left(\mathbf{X}^n, \theta(\mu_0)\right)$ by the bootstrap distribution of $R_n(\mathbf{X}^{*n}, \theta(\widehat{\mu}))$, where $\mathbf{X}^{*n}$ is an IID random sample of size n drawn with replacement from the original sample $\mathbf{X}^n$ using F_n. We use $\widehat{\mu}^*$ to denote the data generating process indexed by the bootstrap empirical distribution $\widehat{F}^*$ defined in analogous way as $\widehat{F}$, that is $\widehat{F}^*(y) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}(X_i^* \leq y)$.

2.1 The single bootstrap

To fix ideas, consider the root function $R_n(\mathbf{X}^n, \theta(\mu_0)) = \theta(\widehat{\mu}) - \theta(\mu_0)$, where $\theta(\widehat{\mu})$ is a *consistent* estimator for $\theta(\mu_0)$. The *theoretical* bias $\beta(\mu_0)$ is defined by the population equation,

$$E_{\mu_0}\left[R_n(\mathbf{X}^n, \theta(\mu_0)) + \beta(\mu_0)\right] = 0. \tag{1}$$

The bootstrap estimate $\beta(\widehat{\mu})$ for the bias correction $\beta(\mu_0)$ is defined by the bootstrap version of (1),

$$E_{\widehat{\mu}}\left[R_n(\mathbf{X}^{*n}, \theta(\widehat{\mu})) + \beta(\widehat{\mu})\right] = 0. \tag{2}$$

Definition 1 *The single bootstrap bias corrected estimator is defined as* $\widehat{\theta}_{bc} = \theta(\widehat{\mu}) + \beta(\widehat{\mu})$, *where* $\beta(\widehat{\mu}) = \theta(\widehat{\mu}) - E_{\widehat{\mu}}\left[\theta(\widehat{\mu}^*)\right]$.

The single bootstrap algorithm. Given the original sample $\mathbf{X}^n$, B bootstrap resamples $\mathbf{X}_b^{*n}, b = 1, \cdots, B$ are randomly drawn from the DGP $\widehat{\mu}$. For each bootstrap resample, the sample value $\widehat{\theta}_b^*$ of the statistic $\theta(\widehat{\mu}^*)$ is computed. A Monte Carlo estimate of $\widehat{\beta}^*$ for the theoretical bias $\beta(\widehat{\mu})$ is calculated using

$$\widehat{\beta}^* \;\; = \;\; \theta(\widehat{\mu}) - \frac{1}{B}\sum_{b=1}^{B}\widehat{\theta}_b^*. \tag{3}$$

The amount of uncorrected bias in $\widehat{\theta}_{bc}$ is of order $O(n^{-2})$; an improvement to the original estimator $\theta(\widehat{\mu})$ which has a bias of order $O(n^{-1})$. For a discussion of the bootstrap refinements, see among others Horowitz (2001), Hall and Horowitz (1996), Efron (1987, 1979) and Efron and Tibshirani (1986).

2.2 The double bootstrap

Beran (1988, 1987) propose the idea of repeated pre-pivoting by mapping a test statistic $\tau_{n,j}$ into a new test statistic $\tau_{n,j+1}$ where $\tau_{n,0}$ is the original sample statistic and $\tau_{n,1}$ is the first bootstrap statistic. The null distribution of $\tau_{n,j}$ is less strongly dependent on the parameters indexing the unknown probability distribution F_0. Hall (1986) shows that the accuracy of the approximation using the j^{th} (iterated) bootstrap critical value is of order $O(n^{-(j+1)/2})$. Furthermore, Shi (1992) shows that the double bootstrap principle can be used without the need of a pivot.

In this section, we use follow Shi (1992), Hall (1992) and Davison and Hinkley (1997) to derive the double bootstrap equation for mean bias correction.

Let $\mathbf{X}^{**n}$ be a (second level) IID random resample of size n drawn with replacement from the first level sample $\mathbf{X}^{*n}$ using F_n. We use $\widehat{\mu}^{**}$ to denote the data generating process indexed by the empirical distribution of $\mathbf{X}^{**n}$.

The likelihood function of $\theta(\widehat{\mu})$ differs from the conditional density function of $\theta(\widehat{\mu}^{*})$, therefore the bootstrap bias estimator $\beta(\widehat{\mu})$ in Definition 1 does not necessarily satisfy the population equation in (1),

$$E_{\mu_0}\left[R_n(\mathbf{X}^n, \theta(\mu_0)) + \beta(\widehat{\mu})\right] \neq 0. \tag{4}$$

Using Davison and Hinkley (1997) notation, to adjust for the deviation from the population equation, a perturbation or adjustment factor is introduced in the form of $b(\widehat{\mu}, \gamma(\mu_0))$ such that,

$$E_{\mu_0}\left[R_n(\mathbf{X}^n, \theta(\mu_0)) + b(\widehat{\mu}, \gamma(\mu_0))\right] = 0. \tag{5}$$

For an additive perturbation, the adjustment takes the form $b(\widehat{\mu}, \gamma(\mu_0)) \equiv \beta(\widehat{\mu}) + \gamma(\mu_0)$.

The bootstrap estimate for $\gamma(\mu_0)$ is generated through the bootstrap version of (5):

$$E_{\widehat{\mu}}\left[R_n(\mathbf{X}^{*n}, \theta(\widehat{\mu})) + b(\widehat{\mu}^{*}, \gamma(\widehat{\mu}))\right] = 0. \tag{6}$$

Notice that $b(\widehat{\mu}^{*}, \gamma(\widehat{\mu}))$ requires a second level bootstrap to estimate $\beta(\widehat{\mu}^{*})$. The bootstrap estimates for $\beta(\widehat{\mu}^{*})$ and $b(\widehat{\mu}^{*}, \gamma(\widehat{\mu}))$ are defined by the sample equations:

$$E_{\widehat{\mu}^{*}}\left[R_n\left(\mathbf{X}^{**n}, \theta(\widehat{\mu}^{*})\right) + \beta(\widehat{\mu}^{*})\right] = 0, \tag{7}$$

$$E_{\widehat{\mu}}\left[R_n(\mathbf{X}^{*n}, \theta(\widehat{\mu})) + b(\widehat{\mu}^{*}, \gamma(\widehat{\mu}))\right] = 0. \tag{8}$$

Combining equations (7) and (8) and assuming an additive adjustment, the double bootstrap estimate of the adjustment factor $\gamma(\widehat{\mu})$ is rewritten as

$$\gamma(\widehat{\mu}) = E_{\widehat{\mu}}\left[\beta(\widehat{\mu}) - \beta(\widehat{\mu}^{*})\right], \tag{9}$$

$$= E_{\widehat{\mu}}\left\{E_{\widehat{\mu}^{*}}\left[\theta(\widehat{\mu}^{**}) - \theta(\widehat{\mu}^{*})\right] - \left[\theta(\widehat{\mu}^{*}) - \theta(\widehat{\mu})\right]\right\}. \tag{10}$$

Definition 2 *The double bootstrap bias estimation $\gamma(\widehat{\mu})$ in equation (9) defines a double bootstrap estimator $\widehat{\theta}_{dbc}$,*

$$\widehat{\theta}_{dbc} = \theta(\widehat{\mu}) + \beta(\widehat{\mu}) + \gamma(\widehat{\mu}).$$

The double bootstrap algorithm. From the original sample $\mathbf{X}^n$, draw B_1 bootstrap resamples $\mathbf{X}_b^{*n}, b = 1, \cdots, B_1$ using the empirical distribution $\widehat{F}_n$. For each resample $\mathbf{X}_b^{*n}$, (i) compute the bootstrap realized value $\widehat{\theta}_b^*$ of $\theta(\widehat{\mu}_b^*)$, (ii) draw B_2 second level bootstrap resamples $\mathbf{X}_{b,j}^{**n}, j = 1, .., B_2$ from the bootstrap empirical distribution $\widehat{F}_b^*$, and (iii) compute the second level bootstrap estimators $\widehat{\theta}_{b,j}^{**}, j = 1, .., B_2$. For $b = 1, \cdots, B_1$, compute an estimate $\widehat{\beta}_b^*$ for the second level bias adjustment $\beta(\widehat{\mu}^*)$ as:

$$\widehat{\beta}_b^* = \widehat{\theta}_b^* - \sum_{j=1}^{B_2} \widehat{\theta}_{b,j}^{**}/B_2.$$

The Monte Carlo estimate of the double bootstrap bias adjustment in equation (9), denoted $\widehat{\gamma}^{**}$, is computed as

$$\widehat{\gamma}^{**} \;\; = \;\; \widehat{\beta} - \frac{1}{B_1} \sum_{b=1}^{B_1} \widehat{\beta}_b^*. \tag{11}$$

The double bootstrap doesn't come cheap. The algorithm makes a total of $B_1(B_2 + 1)$ ($= 249500$ for $B_2 = B_1 = 499$) visits to the statistic $R(\mathbf{X}^{*n}, \theta(\widehat{\mu}))$. This indeed becomes quickly computationally cumbersome depending on the model and the estimation method despite the increase in computational power.

3 Fast methods for approximating the P value

Let us consider the case of estimating the rejection probability and P value of bootstrap tests. Using the notation of Giacomini et al. (2007), consider the root function $R_n(\mathbf{X}^n, \theta(\mu_0))$ with sampling distribution $J_n(\cdot, F)$ and limiting distribution $J(\cdot, F)$. The bootstrap principal approximates the limiting distribution $J(x, F)$ using the bootstrap distribution $\widehat{J}_n(x, F_n^*) \equiv P_n^* \left\{ R_n(\mathbf{X}_k^{*n}, \theta(\widehat{\mu}_k)) \leq x \right\}$.

Standard Monte Carlo experiment.

For each Monte Carlo sample $\mathbf{X}_k^n$ (with DGP $\widehat{\mu}_k$), draw B IID bootstrap samples from $\widehat{\mu}_k$. The bootstrap estimate for $\widehat{J}_{n,k}(\cdot, F_n^*)$ is computed using $J_{n,B,k}^*(x) \equiv B^{-1} \sum_{b=1}^{B} R_n(\mathbf{X}_k^{n*}, \theta(\widehat{\mu}_k)) \leq x$. The quantile $q_{n,B,k}^*$ is then computed by inverting $J_{n,B,k}^*(\cdot)$. For a *left-tail* bootstrap test, the empirical rejection probability is approximated using

$$RP_{n,B,K} = K^{-1} \sum_{k=1}^{K} \mathbf{1} \left\{ R_n(\mathbf{X}_k^n, \theta(\mu_0)) \leq q_{n,B,k}^*(\alpha) \right\}. \tag{12}$$

The standard Monte Carlo method requires $B{\cdot}K$ computations of the root function $R_n(\mathbf{X}^{*n}, \theta(\widehat{\mu}))$. Applying the law of large numbers, $J_{n,B,k}^*$ converges to $\widehat{J}_{n,k}(\cdot, F_n^*)$. Let $\widehat{q}^*(\alpha)$ be the α quantile of $\widehat{J}_n(\cdot, F_n^*)$, $\widehat{J}_n(\widehat{q}^*(\alpha), F_n^*) = \alpha$, then under some regularity conditions (see for example

Giacomini et al. (2007)), $RP_{n,B,K}$ converges to

$$RP = P_{\widehat{\mu}}\left\{R_n(\mathbf{X}^n, \theta(\mu_0)) \le \widehat{q}^*(\alpha)\right\}. \tag{13}$$

Note that the rejection probability in equation (12) also corresponds to the bootstrap P value. This can be seen by rewriting (13) as,

$$RP = P_{\widehat{\mu}}\left\{P_{\widehat{\mu}^*}\left[R_n(\mathbf{X}^{**n}, \theta(\widehat{\mu}^*)) \le R_n(\mathbf{X}^n, \theta(\mu_0))\right] \le \alpha\right\} \tag{14}$$

Warp/fast algorithm.

For each Monte Carlo sample $\mathbf{X}_k^n$, $k = 1, \cdots, K$, draw $B = 1$ bootstrap resample $\mathbf{X}_k^{*n}$ and compute the root $R_n(\mathbf{X}_k^{n*}, \theta(\widehat{\mu}_k))$. Giacomini et al. (2007) provide conditions under which the distribution function $\widehat{J}_n(x, F_n^*)$ can be approximated using,

$$\overline{J}_{n,K}(x) \equiv K^{-1}\sum_{k=1}^{K} \mathbf{1}\left\{R_n(\mathbf{X}_k^{*n}, \theta(\widehat{\mu}_k)) \le x\right\}.$$

An approximation for the bootstrap probability (at nominal level α) follows,

$$\widehat{RP}_{n,K} = K^{-1}\sum_{k=1}^{K} \mathbf{1}\left\{R_n(\mathbf{X}_k^n, \theta(\mu_0)) \le \widehat{q}_{n,K}(\alpha)\right\}, \tag{15}$$

where $\widehat{q}_{n,K}(\alpha)$ is the α quantile of $\overline{J}_{n,K}$ satisfying, $\widehat{q}_{n,K}(\alpha) \equiv inf\{x, \overline{J}_{n,K}(x) \ge \alpha\}$. This is the same estimate $\widehat{RP}_A$ in Davidson and MacKinnon (2007).

The double bootstrap P value (see for example Shi (1992), Davidson and MacKinnon (2002b,a)) is defined as,

$$\widehat{p}^{**} = P_{\widehat{\mu}}\left\{P_{\widehat{\mu}^*}\left[R_n(\mathbf{X}_k^{**n}, \theta(\widehat{\mu}^*)) \le R_n(\mathbf{X}_k^{*n}, \theta(\widehat{\mu}))\right] \le p^*\right\}, \tag{16}$$

where p^* is the first level bootstrap approximation for the P value,

$$p^* = P_{\widehat{\mu}}\left[R_n(\mathbf{X}_k^{*n}, \theta(\widehat{\mu})) \le R_n(\mathbf{X}_k^n, \theta(\mu_0))\right].$$

The fast double bootstrap approximation of Davidson and MacKinnon (2007) is obtained by taking the bootstrap version of $\widehat{RP}_{n,K}$,

$$\widehat{p}^{**}_{n,B} = B^{-1}\sum_{j=1}^{B} \mathbf{1}\left\{R_n(\mathbf{X}_k^{*n}, \theta(\widehat{\mu})) \le \widehat{q}^{**}_{n,B}(\widehat{p}^*)\right\}, \tag{17}$$

where $\widehat{q}^{**}_{n,B}(\alpha) \equiv inf\{x, \overline{J}^*_{n,B}(x) \ge \alpha\}$, $\overline{J}^*_{n,B}(x) \equiv B^{-1}\sum_{b=1}^{B} \mathbf{1}\left\{R_n(\mathbf{X}_b^{**n}, \theta(\widehat{\mu}_b^*)) \le x\right\}$ and $\widehat{p}^* \equiv B^{-1}\sum_{b=1}^{B} \mathbf{1}\left\{R_n(\mathbf{X}_b^{*n}, \theta(\widehat{\mu}_b^*)) \le R_n(\mathbf{X}^n, \theta(\mu_0))\right\}$.

4 Fast approximation of the bootstrap bias correction

In this section we propose a computationally efficient approximation of the calibrating coefficient $\gamma(\widehat{\mu})$ of Definition 2 using the *fast/warp-speed* method described earlier. The double bootstrap adjustment factor $\gamma(\widehat{\mu})$ in equation (10) can be expressed as

$$\gamma(\widehat{\mu}) \quad = \quad E_{\widehat{\mu}}\left\{E_{\widehat{\mu}^*}\left[R_n(\mathbf{X}^{**n}, \theta(\widehat{\mu}_*)) - R_n(\mathbf{X}^{*n}, \theta(\widehat{\mu}))\right]\right\}, \tag{18}$$

where $R_n(\mathbf{X}^n, \theta(\mu_0)) = \theta(\widehat{\mu}) - \theta(\mu_0)$. Notice the analogy between the population equations in (14) and (18). The former is calculating "global probability" and the latter "global expectation". Our proposed approximation of the double bootstrap bias adjustment builds up on this analogy and the results already established for the quantile and the P value functions.

Definition 3 *Let n be a given sample size and B_1 a finite integer. For each first level bootstrap resample $\mathbf{X}_b^{*n}, b = 1, \cdots, B_1$, draw $B_2 = 1$ second level bootstrap resample $\mathbf{X}_b^{**n}$. Let $\overline{R}_{n,B_1}(x) = B_1^{-1}\sum_{b=1}^{B_1}\left[R_n(\mathbf{X}_b^{*n}, \theta(\widehat{\mu})) + x\right]$ and $\overline{R}^*_{n,B_1}(x) = B_1^{-1}\sum_{b=1}^{B_1}\left[R_n(\mathbf{X}_b^{**n}, \theta(\widehat{\mu}^*)) + x\right]$. The first level bootstrap bias $\beta(\widehat{\mu})$ satisfies $\overline{R}_{n,B_1}(\beta(\widehat{\mu})) = 0$. We propose an approximation for $\gamma(\widehat{\mu})$, denoted γ_{FDA}, such that $\overline{R}^*_{n,B_1}(\beta(\widehat{\mu})) = \gamma_{FDA}$: $\gamma_{FDA} = B_1^{-1}\sum_{b=1}^{B_1} R_n(\mathbf{X}_b^{**n}, \theta(\widehat{\mu}^*)) + \beta(\widehat{\mu})$.*
This approximation defines a fast *double bootstrap bias corrected estimator $\widehat{\theta}_{FDA}$*

$$\widehat{\theta}_{FDA} \quad = \quad \theta(\widehat{\mu}) + \beta(\widehat{\mu}) + \gamma_{FDA}. \tag{19}$$

Assumption 1 *$E_{\widehat{\mu}}\left\{R_n(\mathbf{X}^{*n}, \theta(\widehat{\mu}))\right\}$ exists and $E_{\widehat{\mu}}\left|R_n(\mathbf{X}^{*n}, \theta(\widehat{\mu}))\right|^2 < \infty$ for $n = 1, 2, \cdots$.*

The convergence of the *warp-speed* approximation to the bootstrap distribution in Giacomini et al. (2007) does not require an asymptotic pivot or differentiability of the root. Indeed for scalar-valued $\theta(\mu_0)$ and IID data, convergence only requires $R_n : \mathbf{X}^n \times \Re \to \Re$ to be measurable for $n = 1, 2, \cdots$.

Corollary 1 *Suppose that Assumption 1 holds and that $R_n : \mathbf{X}^n \times \Re \to \Re$ is measurable for $n = 1, 2, \cdots$. Then for each n and x,*

$$\overline{R}^*_{n,B_1}(x) \to E_{\widehat{\mu}}E_{\widehat{\mu}^*}\left[R_n(\mathbf{X}^{**n}, \theta(\widehat{\mu}^*)) + x\right] \ \text{as } B_1 \to \infty,$$

and γ_{FDA} converges in probability to $\gamma(\widehat{\mu})$: $\gamma_{FDA} \to \gamma(\widehat{\mu})$.

Proof in Appendix A.

Implementing the fast double bootstrap approximation. For each first level bootstrap resample $\mathbf{X}_b^{*n}$: (i) compute $\widehat{\theta}_b^*$, (ii) draw $B_2 = 1$ second level bootstrap resample $\mathbf{X}_b^{**n}$ and, (iii) compute $\widehat{\theta}_b^{**}$. After all bootstrapping operations are complete, we have two series of bootstrap iterates, $\widehat{\theta}_b^*$ and $\widehat{\theta}_b^{**}$ for $b = 1, \cdots, B_1$. The first level bootstrap bias $\beta(\widehat{\mu})$ is estimated in a way similar to that given in (3). An estimate $\widehat{\gamma}_{DFA}$ of the proposed fast double bootstrap approximation γ_{DFA} is computed as

$$\widehat{\gamma}_{FDA} \quad = \quad \widehat{\beta} + \frac{1}{B_1}\sum_{b=1}^{B_1}\left(\widehat{\theta}_b^{**} - \widehat{\theta}_b^*\right). \tag{20}$$

This algorithm requires only $2B_1 + 1$ visits to the statistic of interest θ_μ. Therefore the computational cost is reduced from order $O(B_1 B_2)$ in the double bootstrap to $O(B_1)$ for the proposed fast approximation.

5 Monte Carlo Analysis

5.1 The Monte Carlo Environment

Consider the simple linear IV model of Guggenberger (2008)

$$y_i = \theta x_i + \epsilon_i, \tag{21}$$

$$x_i = z_i \pi + v_i \qquad i = 1, ..., n. \tag{22}$$

For simplicity, we assume that the endogenous variable in the left-hand side of (21) is a scalar. The scalar regressor x_i is endogenous and accepts the reduced form in (22). The $K - vector$ z_i represents the predetermined/exogenous instruments which satisfies exogeneity condition, $E(z_i \epsilon_i) = 0$. The random variables z_i is IID normally distributed random variables $N(0, I_K)$, and (ϵ, v_i) are IID $N(0, \Omega)$ with $\Omega = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$.

Two parameters are of special interest in this model and will affect the bias of the IV estimator. First, the correlation parameter ρ which determines the degree of endogeneity of x_i. Secondly, the strength of the instruments π which measure the relevance of the instruments. If the latter is zero, the IV estimator is neither consistent nor asymptotically normal. To control for this parameter we use the $\mathbb{R}^2$ from the first stage regression which is equal to, $R^2 = \frac{\pi'\pi}{1+\pi'\pi}$.

Assuming that all the instruments have the same strength η (See, Guggenberger (2008)) or alternatively if the total explanatory power of the first stage regression is equally assigned among $\pi_j = \eta, j = 1 : K$ (Flores-Lagunes (2007)), the R^2 is thus related to the relevance of the instruments and to the number of instruments in the simple equation $R^2 = \frac{K.\eta^2}{1+K.\eta^2}$. The IV estimator $\widehat{\theta} = (\mathbf{x}' P_z \mathbf{x})^{-1} \mathbf{x}' P_z \mathbf{y}$, where $P_z = \mathbf{z}(\mathbf{z}'\mathbf{z})^{-1}\mathbf{z}'$, is consistent. The finite sample bias of $\widehat{\theta}$ is dependent on ρ, π and K as follows, $bias(\widehat{\theta}) = \frac{(K-2).\rho}{n.(\pi'\mathbf{z}'\mathbf{z}\pi)^{-1}}$. The data are simulated to represent cases of weak or less relevant instruments (low R^2) and cases of high endogeneity of x_i (high ρ). The degree of overidentification (number of instruments K) also plays a role in the tradeoff between bias and efficiency for IV estimation. We therefore consider experiments with the following combinations of the in the DGP : $n \in \{50, 200\}$, $R^2 \in \{0.01, 0.15, 0.25\}$, $K \in \{5, 10\}$ and $\rho \in \{0.25, 0.50, 0.85\}$. In all experiments, the true DGP μ_0 is characterized by $\theta = 0$.

5.2 Monte Carlo Results

Because of the increased computational time due to the nested sampling, we limit the number of Monte Carlo simulations to $10,000$ and the number of bootstrap iterations to $B_1 = B_2 = 499$ (considered as reasonable in Davidson and MacKinnon (2007)).

Tables I-III report the summary statistics of the empirical distributions of the instrumental variable estimator $\widehat{\theta}$, the bootstrap bias corrected estimator $\widehat{\theta}_{bc}$, the double bootstrap bias corrected estimator $\widehat{\theta}_{dbc}$, and the bias corrected estimator using the fast approximation $\widehat{\theta}_{FDA}$. In particular, the columns show the sample mean, the sample median (Med), the root mean square error RMSE, and the mean absolute error MAE. Figure 1 plots the kernel density estimates of the empirical distributions of the bootstrap estimators for selected DGP parametrization.

The results for the linear IV estimator $\widehat{\theta}$ are conventional. As the degree of overidentification increases with the number od instruments moving from $K = 5$ to $K = 10$, there is an increase in the mean and median bias while the RMSE goes down. Increasing the sample size results in considerable decrease in the bias and increase in efficiency as can be seen when comparing Table II to Table III. If the instruments are strong (high R^2), the presence of endogeneity is not a disaster even for high values of ρ. In Table I for example, the bias of $\widehat{\theta}$ increased from 0.04 to 0.07 when ρ increased from 0.25 to 0.85

The statistical properties of the single bias corrected estimator $\widehat{\theta}_{bc}$ and the double bootstrap bias corrected estimator $\widehat{\theta}_{dbc}$ in tables I-III follow existing predictions in the literature. For all the DGP configurations with $R^2 \in \{0.15, 0.25\}$, the bias of $\widehat{\theta}_{dbc}$ is smaller than that of $\widehat{\theta}_{bc}$ which in turn is significantly reduced compared to $\widehat{\theta}$. In Table II for $R^2 = 0.25$ and $\rho = 0.85$, the single bootstrap reduces the bias by 29% while the double bootstrap further reduces the gap by 59%. These percentage reductions are higher (41% and 85% respectively) when $\rho = 0.25$. However, reducing the bias may increase the variance, or even the mean squared error. Indeed, this is the case in most configurations except for the case of $R^2 = 0.25$ and $\rho = 0.85$ and $K = 10$ where the RMSE of $\widehat{\theta}$ decreased by 3.8% for $n = 50$.

This result is not new. Hsu et al. (1986) find that the bias reduction from bootstrapping the two stage least squares is achieved at the expense if increased variance. MacKinnon and Smith (1998) argue that reducing the bias may increase the variance and the root mean squared error of the bias corrected estimator depending on the shape of the bias function and on the variance of the initial estimator. We conjecture that the shape of the bias function of the IV estimator depends on the relevance of the instruments and the severity of the endogeneity problem. In addition, the increased variance may be due to few erratic bootstrap estimates $\widehat{\theta}_b^*$ and $\widehat{\theta}_{b,j}^{**}$. Following Hsu et al. (1986) and Shao (1990), we compute the adjusted root mean squared error RMSE$_a$ by deleting 2.5% from the top and 2.5% from the bottom of the $5,000$ bias corrected estimates. The results show no significant increase in the adjusted RMSE for $\widehat{\theta}_{bc}$. The mean bias of the double bootstrap estimator $\widehat{\theta}_{dbc}$ is smaller than that of the single bootstrap estimator $\widehat{\theta}_{bc}$. However, this gain in bias does not offset the increased RMSE and RMSE$_a$.

The proposed approximation of the double bootstrap bias produces estimators with higher bias reduction than the single bootstrap for all configurations. This reduction comes with lower variance than the double bootstrap corrected estimator. The proposed estimator $\widehat{\theta}_{FDA}$ has lower mean absolute error MAE and RMSE than $\widehat{\theta}_{dbc}$. In addition to reduced bias, this estimator has lower RMSE$_a$ and MAE than the IV estimator $\widehat{\theta}$ and the single bootstrap bias corrected estimator $\widehat{\theta}_{bc}$. See for example the underlined results in tables II-III.

The results are even more promising for the proposed approximation when the sample size is increased from $n = 50$ to $n = 200$. Table III shows that for models with relevant instruments, the proposed bias corrected estimator not only achieves a lower bias than the single bootstrap but also outperforms the double bootstrap corrected estimator. In Table III when $\mathbb{R}^2 = 0.25$ and $\rho = 0.85$, $\widehat{\theta}_{dbc}$ reduces the bias of $\widehat{\theta}$ from 0.039 to -0.019 while $\widehat{\theta}_{FDA}$ further reduces the bias to 0.003.

Figure 1 provides an overview of the effect of instrument relevance and sample size on the sampling distributions of the alternative estimators. The plots show that in the presence of weak instruments (plots with $\mathbb{R}^2 = 0.01$), the bias of all estimators is higher, the increase in the variance of the bootstrap bias corrected estimators is also higher. Increasing the sample size from $n = 50$ to $n = 200$ has little effect on both the bias and the variance unless the instruments are relevant and the endogeneity is moderate (see Figure 1(c)). In addition, the evidence from the Monte Carlo suggests that the bootstrap is less effective in reducing the finite sample bias when the instruments are weak.

6 Conclusion

The theory predicts that iterating the bootstrap principle increases the accuracy of the bootstrap. This increased accuracy comes at an enormous computational cost. This paper has presented a new computationally efficient technique for bias correction which removes the requirement to perform the computationally intensive inner loop of the double bootstrap. The proposed approximation converges to the theoretical bias adjustment of the double bootstrap. The new bootstrap bias corrected estimator is effective in reducing the bias of the single bootstrap and, in the example considered, is more precise than the double bootstrap. In the linear instrumental variable model, the bias function of the IV estimator depends on the instruments relevance, the degree of endogeneity and the number of instruments. We find that the bootstrap as a method to reduce the bias is less effective when the instruments are weak regardless of the sample size. In the case of weak identification, models with high degree of endogeneity have lower mean and median bias. This result warrants further investigation.

References

Beran, R. (1990) “Refining Bootstrap Simultaneous Confidence Sets,” *Journal of Statistical Planning and Inference* **43**, 205–213.

Beran, R. (1988) “Prepivoting Test Statistics: a Bootstrap View of Asymptotic Refinements,” *Journal of the American Statistical Association* **83**, 687–697.

Beran, R. (1987) “Prepivoting to Reduce Level Error in Confidence Sets,” *Biometrika* **74**, 457–468.

Davidson, R. and J. G. MacKinnon (2007) “Improving the Reliability of Bootstrap Tests with the Fast Double Bootstrap,” *Computational Statistics and Data Analysis* **51** (7), 3259–3281.

Davidson, R. and J. G. MacKinnon (2002a) “Bootstrap J tests of Nonnested Linear Regression Models,” *Journal of Econometrics* **109**, 167–193.

Davidson, R. and J. G. MacKinnon (2002b) “Fast Double Bootstrap Tests of Nonnested Linear Regression Models,” *Econometric Reviews* **21**, 417–427.

Davison, C., A and D. V. Hinkley (1997) “Bootstrap methods and their application,” *In Cambridge Series on Statistical and Probabilistic Mathematics* (Cambridge University Press).

Efron, B. (1987) “Better Bootstrap Confidence Intervals,” *Journal od American Statistical Association* **82**, 171–200.

Efron, B. (1979) “Bootstrap Methods: Another Look at the Jackknife,” *Annals of Statistics* **7**, 1–26.

Efron, B. and R. Tibshirani (1986) “Bootstrap Methods for Standard Errors, Confidence Intervals, and other Measures of Statistical Accuracy,” *Statistical Science* **1** (1), 54–77.

Flores-Lagunes, A. (2007) “Finite Sample Evidence of IV Estimators under Weak Instruments,” *Journal of Applied Econometrics* **22**, 677–694.

Giacomini, R., D. N. Politis, and H. White (2007) “A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators,” *Unpublished manuscript.*

Guggenberger, P. (2008) “Finite Sample Evidence Suggesting a Heavy Tail Problem of the Genralized Method of Empirical Likelihood Estimator,” *Econometric Reviews* **26**, 526–541.

Hall, P. (1992) *The Bootstrap and Edgeworth Expansion*, New York: Springer-Verlag.

Hall, P. (1986) “On the Bootstrap and Confidence Intervals,” *The Annals of Statistics* **14** (4), 1431–1452.

Hall, P. and L. Horowitz, Joel (1996) “Bootstrap Critical Values for Tests Based on Generalized Method of Moments Estimators,” *Econometrica* **64**, 891–916.

Hall, P. and M. A. Martin (1988) “On Bootstrap Resampling and Iteration,” *Biometrika* **75**, 661–671.

Horowitz, J. L. (2001) *The Bootstrap* **5**, Chap. 52, 3159–3228: Heckman, J.J. and Leamer, E.E. (eds), Elsevier Science. in Handbook of Econometrics.

Hsu, Y. S., K. N. Lau, H. G. Fung, and E. F. Ulveling (1986) “Monte Carlo studies on the effectiveness of the Bootstrap bias reduction method on 2SLS estimates,” *Economics Letters*, 233–239.

Lee, S. M. S. and G. A. Young (1999) “The effect of Monte Carlo approximation on coverage error of double-bootstrap confidence intervals,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **61** (2), 353–366.

MacKinnon, J. G. and A. A. Smith (1998) “Approximate bias correction in econometrics,” *Journal of Econometrics* (2), 205–230.

Shao, J. (1990) “Bootstrap estimation of the asymptotic variances of statistical functionals,” *Annals of the Institute of Statistics and Mathematics* (4), 737–752.

Shi, S. G. (1992) “Accurate and Efficient Double-bootstrap Confidence Limit Method,” *Computational Statistics and Data Analysis*, 21–32.

Appendix A: Proof of Corollary 1

We first establish convergence of $\overline{R}^*_{n,B_1}(x)$ to $E_{\widehat{\mu}}E_{\widehat{\mu}^*}\left[R_n(\mathbf{X}^{**n},\theta(\widehat{\mu}_*))+x\right]$:

$$\overline{R}^*_{n,B_1}(x) \quad = \quad B_1^{-1}\sum_{b=1}^{B_1} R_n(\mathbf{X}^{**n}_{l,b},\theta(\widehat{\mu}^*_b))+x \tag{23}$$

Note that x depends only on the DGP $\widehat{\mu}$ and therefore can be taken out of the conditional expectation. By the assumption of IID random draws,

$$E_{\widehat{\mu}^*}\overline{R}^*_{n,B_1}(x) = E_{\widehat{\mu}^*}\left\{R_n(\mathbf{X}^{**n}_b,\theta(\widehat{\mu}^*_b))\right\}+x,$$

and the variance

$$V_{\widehat{\mu}^*}\overline{R}^*_{n,B_1}(x) = B_1^{-1}V_{\widehat{\mu}^*}R_n(\mathbf{X}^{**n}_b,\theta(\widehat{\mu}^*_b)) \to 0, \text{ as } B_1\to\infty.$$

Applying the law of large numbers ,

$$B_1^{-1}\sum_{l=1}^{B_1} R_n(\mathbf{X}^{**n}_{l,b},\theta(\widehat{\mu}_*)) \to E_{\widehat{\mu}^*}R_n(\mathbf{X}^{**n}_{l,b},\theta(\widehat{\mu}^*_b)) \equiv E_{\widehat{\mu}^*}R_n(\mathbf{X}^{**n}_b,\theta(\widehat{\mu}^*_b)) \tag{24}$$

as $B_1 \to \infty$. Similarly, using the law of large numbers and given that $E_{\widehat{\mu}^*} R_n(\mathbf{X}_b^{**n}, \theta(\widehat{\mu}_b^*))$ are IID random variables,

$$\begin{aligned} E_{\widehat{\mu}} \overline{R}^*_{n,B_1}(x) &= E_{\widehat{\mu}} E_{\widehat{\mu}^*} \overline{R}^*_{n,B_1}(x) \\ &= B_1^{-1} \sum_{b=1}^{B_1} E_{\widehat{\mu}} E_{\widehat{\mu}^*} R_n(\mathbf{X}_b^{**n}, \theta(\widehat{\mu}_b^*)) + x \\ &= E_{\widehat{\mu}} E_{\widehat{\mu}^*} \left[R_n(\mathbf{X}^{**n}, \theta(\widehat{\mu}_*)) + x \right]. \end{aligned}$$

In addition, given the IID assumption and

$$V_{\widehat{\mu}} \overline{R}^*_{n,B_1}(x) = B_1^{-1} V_{\widehat{\mu}} E_{\widehat{\mu}^*} R_n(\mathbf{X}^{**n}, \theta(\widehat{\mu}^*)) < \infty,$$

therefore using the law of large numbers we establish, $\overline{R}^*_{n,B_1}(x) \to E_{\widehat{\mu}} E_{\widehat{\mu}^*} R_n(\mathbf{X}^{**n}, \theta(\widehat{\mu}^*)) + x$ as $B_1 \to \infty$. Similar result can be established for $\overline{R}_{n,B_1}(x)$ by using the law of large numbers and the fact that the random variables $R_n(\mathbf{X}_b^{*n}, \theta(\widehat{\mu}_b))$ are IID:

$$\overline{R}_{n,B_1}(x) \to E_{\widehat{\mu}} \left[R_n(\mathbf{X}^{*n}, \theta(\widehat{\mu}^*)) + x \right], \text{ as } B_1 \to \infty.$$

The fast approximation of the adjustment parameter γ_{FDA} satisfies $\gamma_{FDA} = \overline{R}^*_{n,B_1}(\beta(\widehat{\mu}))$. Solving for γ_{FDA},

$$\begin{aligned} \gamma_{FDA} &= \frac{1}{B_1} \sum_{b=1}^{B_1} (\theta(\widehat{\mu}_b^{**}) - \theta(\widehat{\mu}_b^*)) + \beta(\widehat{\mu}) & (25) \\ &= \left[\theta(\widehat{\mu}) - \frac{1}{B_1} \sum_{b=1}^{B_1} \theta(\widehat{\mu}_b^*) \right] + \left[\frac{1}{B_1} \sum_{b=1}^{B_1} (\theta(\widehat{\mu}_b^{**}) - \theta(\widehat{\mu}_b^*)) \right] & (26) \\ &= \left[\frac{1}{B_1} \sum_{b=1}^{B_1} (\theta(\widehat{\mu}_b^{**}) - \theta(\widehat{\mu}_b^*)) \right] - \left[\frac{1}{B_1} \sum_{b=1}^{B_1} (\theta(\widehat{\mu}_b^*) - \theta(\widehat{\mu})) \right] & (27) \\ &= \overline{R}^*_{n,B_1}(0) - \overline{R}_{n,B_1}(0) & (28) \\ &\to E_{\widehat{\mu}} E_{\widehat{\mu}^*} \left[R_n(\mathbf{X}^{**n}, \theta(\widehat{\mu}_*)) \right] - E_{\widehat{\mu}} \left[R_n(\mathbf{X}^{*n}, \theta(\widehat{\mu}^*)) \right] & (29) \\ &\to \gamma(\widehat{\mu}), \text{ as } B_1 \to \infty. & (30) \end{aligned}$$

Table I: Monte Carlo results (10,000 replications; true $\theta = 0$; number of instruments $K = 5$; sample size $n = 50$). Summary statistics for the distribution of alternative estimators of θ: the mean, the median (Med), root mean square error (RMSE), adjusted RMSE (RMSE$_a$) and the mean absolute deviation (MAE).

	Mean	Med	RMSE	RMSE$_a$	MAE	Mean	Med	RMSE	RMSE$_a$	MAE	Mean	Med	RMSE	RMSE$_a$	MAE
	$R^2 = 0.25, \rho = 0.85$					$R^2 = 0.25, \rho = 0.50$					$R^2 = 0.25, \rho = 0.25$				
$\widehat{\theta}$	0.07	0.07	0.12	0.12	0.10	0.07	0.07	0.20	0.20	0.16	0.04	0.04	0.22	0.22	0.18
$\widehat{\theta}_{bc}$	0.026	0.043	0.14	0.11	0.11	0.02	0.04	0.23	0.19	0.18	0.00	0.02	0.27	0.21	0.20
$\widehat{\theta}_{dbc}$	-0.02	0.01	0.18	0.13	0.12	-0.03	-0.00	0.29	0.22	0.21	-0.02	-0.01	0.32	0.25	0.24
$\widehat{\theta}_{FDA}$	0.01	0.04	0.16	0.12	0.11	0.00	0.03	0.26	0.20	0.18	0.00	0.01	0.28	0.21	0.20
	Mean	Med	RMSE	RMSE$_a$	MAE	Mean	Med	RMSE	RMSE$_a$	MAE	Mean	Med	RMSE	RMSE$_a$	MAE
	$R^2 = 0.15, \rho = 0.85$					$R^2 = 0.15, \rho = 0.50$					$R^2 = 0.15, \rho = 0.25$				
$\widehat{\theta}$	0.13	0.14	0.19	0.19	0.16	0.11	0.12	0.25	0.25	0.20	0.07	0.07	0.30	0.30	0.23
$\widehat{\theta}_{bc}$	0.09	0.11	0.23	0.18	0.17	0.05	0.07	0.31	0.24	0.23	0.03	0.04	0.41	0.29	0.28
$\widehat{\theta}_{dbc}$	0.04	0.08	0.29	0.20	0.20	-0.01	0.03	0.40	0.29	0.28	-0.01	0.02	0.52	0.35	0.34
$\widehat{\theta}_{FDA}$	0.06	0.09	0.28	0.19	0.19	0.02	0.06	0.36	0.26	0.25	0.02	0.04	0.48	0.37	0.31
	Mean	Med	RMSE	RMSE$_a$	MAE	Mean	Med	RMSE	RMSE$_a$	MAE	Mean	Med	RMSE	RMSE$_a$	MAE
	$R^2 = 0.01, \rho = 0.85$					$R^2 = 0.01, \rho = 0.50$					$R^2 = 0.01, \rho = 0.25$				
$\widehat{\theta}$	0.25	0.25	0.30	0.30	0.26	0.33	0.33	0.50	0.50	0.40	0.23	0.23	0.55	0.55	0.42
$\widehat{\theta}_{bc}$	0.24	0.24	0.37	0.31	0.30	0.32	0.32	0.69	0.54	0.51	0.23	0.23	0.84	0.63	0.60
$\widehat{\theta}_{dbc}$	0.24	0.24	0.46	0.36	0.34	0.32	0.31	0.92	0.66	0.64	0.24	0.23	1.14	0.82	0.79
$\widehat{\theta}_{FDA}$	0.24	0.24	0.44	0.34	0.33	0.32	0.32	0.87	0.62	0.60	0.24	0.23	1.07	0.76	0.73

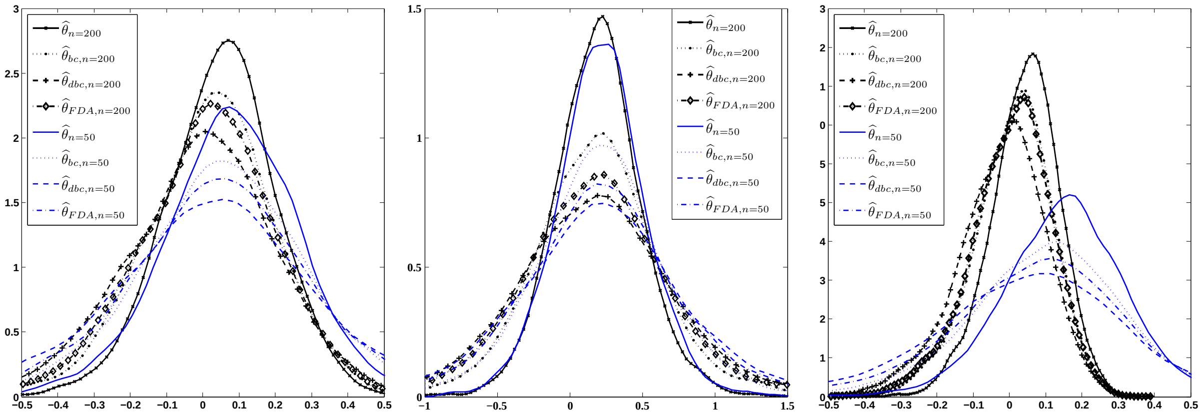
Table II: Monte Carlo results (10,000 replications; true $\theta = 0$; number of instruments $K = 10$; sample size $n = 50$).

	Mean	Med	RMSE	RMSE$_a$	MAE	Mean	Med	RMSE	RMSE$_a$	MAE	Mean	Med	RMSE	RMSE$_a$	MAE
	$R^2 = 0.25, \rho = 0.85$					$R^2 = 0.25, \rho = 0.50$					$R^2 = 0.25, \rho = 0.25$				
$\widehat{\theta}$	0.13	0.13	0.16	0.16	0.14	0.15	0.15	0.22	0.22	0.18	0.08	0.08	0.21	0.21	0.16
$\widehat{\theta}_{bc}$	0.09	0.10	0.15	0.13	0.12	0.09	0.11	0.24	0.20	0.19	0.05	0.05	0.25	0.20	0.19
$\widehat{\theta}_{dbc}$	0.05	0.07	0.16	0.13	0.13	0.04	0.06	0.29	0.23	0.22	0.01	0.03	0.30	0.24	0.23
$\widehat{\theta}_{FDA}$	0.07	0.08	0.15	0.12	0.12	0.07	0.09	0.27	0.23	0.20	0.03	0.04	0.27	0.23	0.21
	Mean	Med	RMSE	RMSE$_a$	MAE	Mean	Med	RMSE	RMSE$_a$	MAE	Mean	Med	RMSE	RMSE$_a$	MAE
	$R^2 = 0.15, \rho = 0.85$					$R^2 = 0.15, \rho = 0.50$					$R^2 = 0.15, \rho = 0.25$				
$\widehat{\theta}$	0.16	0.16	0.18	0.18	0.16	0.20	0.21	0.28	0.28	0.23	0.11	0.11	0.25	0.25	0.20
$\widehat{\theta}_{bc}$	0.12	0.13	0.18	0.16	0.15	0.15	0.16	0.31	0.27	0.25	0.07	0.07	0.31	0.25	0.24
$\widehat{\theta}_{dbc}$	0.09	0.10	0.20	0.16	0.15	0.10	0.12	0.37	0.30	0.29	0.04	0.04	0.39	0.31	0.29
$\widehat{\theta}_{FDA}$	0.10	0.11	0.19	0.16	0.15	0.13	0.14	0.34	0.28	0.27	0.05	0.05	0.35	0.28	0.27
	Mean	Med	RMSE	RMSE$_a$	MAE	Mean	Med	RMSE	RMSE$_a$	MAE	Mean	Med	RMSE	RMSE$_a$	MAE
	$R^2 = 0.01, \rho = 0.85$					$R^2 = 0.01, \rho = 0.50$					$R^2 = 0.01, \rho = 0.25$				
$\widehat{\theta}$	0.26	0.25	0.28	0.28	0.26	0.33	0.33	0.41	0.41	0.35	0.22	0.22	0.38	0.38	0.30
$\widehat{\theta}_{bc}$	0.25	0.25	0.30	0.29	0.26	0.33	0.33	0.49	0.44	0.40	0.21	0.21	0.52	0.43	0.40
$\widehat{\theta}_{dbc}$	0.25	0.25	0.34	0.31	0.28	0.32	0.33	0.60	0.50	0.47	0.21	0.21	0.67	0.54	0.51
$\widehat{\theta}_{FDA}$	0.25	0.25	0.32	0.30	0.28	0.32	0.33	0.57	0.48	0.45	0.21	0.21	0.63	0.50	0.47

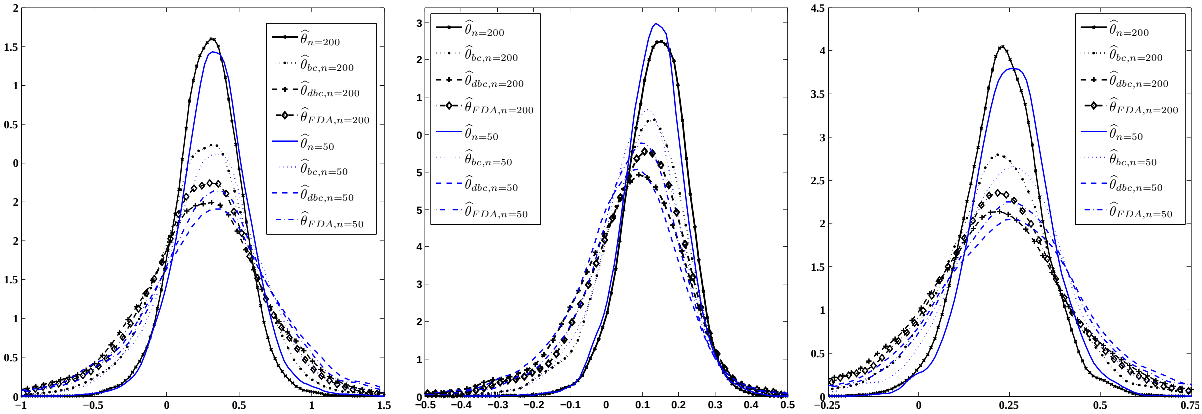
Table III: Monte Carlo results (10,000 replications; true $\theta = 0$; number of instruments $K = 10$; sample size $n = 200$).

	Mean	Med	RMSE	RMSE$_a$	MAE	Mean	Med	RMSE	RMSE$_a$	MAE	Mean	Med	RMSE	RMSE$_a$	MAE
	$R^2 = 0.25, \rho = 0.85$					$R^2 = 0.25, \rho = 0.50$					$R^2 = 0.25, \rho = 0.25$				
$\widehat{\theta}_{IV}$	0.04	0.04	0.06	0.06	0.05	0.04	0.05	0.10	0.10	0.08	0.05	0.05	0.16	0.16	0.13
$\widehat{\theta}_{bc}$	0.01	0.02	0.06	0.05	0.05	0.01	0.02	0.11	0.09	0.08	0.02	0.03	0.18	0.15	0.14
$\widehat{\theta}_{dbc}$	-0.02	-0.01	0.08	0.07	0.07	-0.02	-0.01	0.12	0.10	0.10	-0.01	0.00	0.22	0.18	0.17
$\widehat{\theta}_{FDA}$	0.00	0.01	0.07	0.06	0.05	0.00	0.01	0.11	0.09	0.09	0.01	0.02	0.20	0.16	0.15
	Mean	Med	RMSE	RMSE$_a$	MAE	Mean	Med	RMSE	RMSE$_a$	MAE	Mean	Med	RMSE	RMSE$_a$	MAE
	$R^2 = 0.01, \rho = 0.85$					$R^2 = 0.01, \rho = 0.50$					$R^2 = 0.01, \rho = 0.25$				
$\widehat{\theta}$	0.23	0.23	0.26	0.26	0.24	0.29	0.30	0.37	0.37	0.32	0.20	0.20	0.35	0.36	0.28
$\widehat{\theta}_{bc}$	0.22	0.22	0.28	0.26	0.24	0.27	0.28	0.43	0.38	0.35	0.18	0.18	0.48	0.40	0.37
$\widehat{\theta}_{dbc}$	0.21	0.21	0.30	0.27	0.25	0.25	0.26	0.52	0.44	0.41	0.17	0.17	0.62	0.50	0.47
$\widehat{\theta}_{FDA}$	0.21	0.22	0.29	0.27	0.25	0.26	0.27	0.49	0.42	0.39	0.18	0.18	0.58	0.46	0.44

Figure 1: Kernel density estimates of the empirical distribution of bootstrap biased corrected estimators of θ. Monte Carlo experiments: 10,000 replications; $\theta = 0$; $)K = 10$; $n =\in \{50, 200\}$; $\rho \in \{0.25, 0.50, 0.85\}$; $R^2 \in \{0.01, 0.25\}$.



(a) $\rho = 0.25, R^2 = 0.25$ (b) $\rho = 0.25, R^2 = 0.01$ (c) $\rho = 0.50, R^2 = 0.25$



(d) $\rho = 0.50, R^2 = 0.01$ (e) $\rho = 0.85, R^2 = 0.25$ (f) $\rho = 0.85, R^2 = 0.01$