Fast Iterated Bootstrap Mean Bias Correction

Rachida Ouysse[1]

School of Economics, The University of New South Wales, Sydney 2052 Australia.

Abstract

Abstract: The article proposes a computationally efficient procedure for bias adjustment in the iterated bootstrap. The new technique replaces the need for successive levels of bootstrap resampling by proposing an approximation for the double bootstrap "calibrating coefficient" using only one draw from the second level probability distribution. Extensive Monte Carlo evidence suggest that the proposed approximation performs better than the ordinary bootstrap bias correction. The article evaluates the usefulness of the bootstrap and fast bootstrap in reducing the bias of generalized method of moments estimators under weak instruments. In identified models, this fast bootstrap bias correction leads to estimators with lower variance than those based on the double bootstrap. The proposed fast iterated bootstrap performs better than the double bootstrap in all scenarios and especially when the model has the weakest instrument relevance and the highest degree of endogeneity. However, when the estimators have no finite moments and the instruments are weak, the bootstrap does not work well and iterating it makes things worse.

Key words: GMM estimation, Bootstrap, double-bootstrap, bias-correction, Monte Carlo simulation, consumption-based asset-pricing model.
JEL Classification: C12;C13;C15

1. Introduction

This paper describes an alternative technique to the double bootstrap for bias correction. This technique uses results from the "fast double bootstrap" procedure introduced by Davidson and MacKinnon [2007] to achieve bias reduction of order higher than the single bootstrap with fraction of the computational cost of the iterated bootstrap.

The idea of iterating the principle of the bootstrap proves to achieve higher order refinements for correction estimation bias and confidence bounds. Beran [1988] argued than pre-pivoting reduces the dependence between the probability distribution of the resample and the unknown data generating process. Therefore, resamples reinforce 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. The higher order refinements of the double/iterated bootstrap

Email address: `rouysse@unsw.edu.au.` (Rachida Ouysse)

appear in many studies, *inter alia* Beran [1988, 1987, 1990], Hall [1992, 1986], Hall and Martin [1988], Lee and Young [2002] and Shi [1992].

In principle, the bootstrap can be iterated to reduce the bias by a factor of $O(n^{-1})$ successively. The general framework presented by Hall and Martin [1988] is based on the idea of calculating an adjustment factor or "calibrating coefficient" to correct the single bootstrap approximation to the quantities of interest such as estimation bias, confidence interval bounds and nominal level.

In practice however, one seldom sees the iterated bootstrap beyond the second level (double bootstrap) due to the increasing computational intensity. Indeed, depending on the model and the estimation method, even the double bootstrap can be computationally very demanding. This has prompted a number of authors to develop computationally efficient and cheaper alternatives to compounded sampling. The common goal of these studies is to find approximations to the iterated bootstrap which eliminate the need for nested levels of resampling.

Davidson and MacKinnon [2007, 2002a] propose a technique they call fast double bootstrap (FDB) to improve the reliability of bootstrap estimate for the error in rejection probability and bootstrap P values while bypassing the computational cost that involves the double bootstrap. They argue that the *FDB* modified P values "will tend to be similar to the ordinary bootstrap P value when the latter is reliable but more accurate when it is unreliable." In other terms, the modified P value is both an improvement over the single bootstrap P value approximation and a check of its accuracy.

In this paper, we demonstrate how we may obtain approximations of the adjustment factor without the use of the full second level bootstrap sampling. Indeed, as in Davidson and MacKinnon [2007], we only require one draw from the second level bootstrap distribution. The number of computations required are only twice what is usually needed to perform the single bootstrap.

We provide Monte Carlo evidence of finite sample properties of bias correction of the proposed method in two examples. The first example is a simple linear instrumental variables model as in Hahn and Hausman [2002] (also in Guggenberger [2008]). The second example is the consumption capital market model (C-CAPM), a leading application of a nonlinear-in parameters generalized method of moments model. The GMM in the second example nests the linear IV model as a special case. Our aim is to evaluate the new proposed efficient bootstrap bias correction technique in relation to the issue of weak instrument and weak identification in these class of models. Research on weak identification in nonlinear models remains mostly based on large-sample approximations and mainly focussed on robust tests and confidence sets. GMM and *IV* estimators are inconsistent and their limiting distributions are nonstandard. It is an open question whether the (iterated) bootstrap is successful in reducing the bias in finite samples. We add to the understanding of the consequences of GMM estimation under weak instruments in two main ways. First, we provide finite sample evidence through Monte Carlo simulation regarding point estimation and bootstrap bias correction in DGPs with varying degrees of instruments relevance. Second, the usefulness and the consequences of the double bootstrap as a method for achieving higher bias-reduction is evaluated. The results provide guidelines on when the bootstrap and the more sophisticated *FDB* can be expected to work well.

The paper is organized as follows. Section two reviews the single and double bootstrap methods for bias estimation. Section three defines the new fast double bootstrap method and discusses its implementation. Section four describes the Monte Carlo environment for the calibrated models in the linear IV and nonlinear GMM . It discusses the computational considerations and simulation algorithms. Section five presents the Monte Carlo results. Section six concludes with discussion of future research.

2. Alternative Bootstraps for Bias Estimation

2.1. Single Bootstrap Bias Correction

Let $\mathbf{X}$ be a random variable with unknown probability distribution function F_0 indexed by a parameter θ_0. Consider a random sample from the data generating process μ_0 of $\mathbf{X}$ with realization $\mathbf{x} = (x_1, x_2, ..., x_n)$. Let $\widehat{\mu}$ be the data generating process determined by $\widehat{F}$, some estimate of the empirical distribution implied by $\mathbf{x}$. A standard uniform choice for is $\widehat{F}(y) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq y)$ and $I(.)$ stands for the indicator function which takes the value of one if the statement in its argument is true. Other choices of $\widehat{F}$ are used to allow nonuniform sampling from the data. The choice of $\widehat{F}$ does not affect the result in this paper.

Let the random variable $R_{t(\mu_0)}(\widehat{\mu}, \mu_0)$ be some deterministic function of $\widehat{\mu}$, μ_0 and some parameter $t(\mu_0)$.

Many statistical problems can be formulated as specifying the statistical properties of the random variable $R_{t(.)}(.,.)$ such as its probability distribution, moments and percentiles. For a given parameter of interest $\theta(\mu_0)$ and a *consistent* estimator $\theta(\widehat{\mu})$, the Root function for the statistical problem of bias estimation is defined as[1]

$$R_{t(\mu_0)}(\widehat{\mu}, \mu_0) \quad = \quad \theta(\widehat{\mu}) - \theta(\mu_0) + t(\mu_0) \tag{1}$$

where the bias $t(\mu_0)$ in estimating $\theta(\mu_0)$ satisfies the population equation

$$E\left\{R_{t(\mu_0)}(\widehat{\mu}, \mu_0)|\mu_0\right\} \quad = \quad 0. \tag{2}$$

The aim is to compute the bias $t(\mu_0)$. If we were able to repeatedly draw samples from μ_0 for a known $\theta(F_0)$, then the *theoretical* bias corresponding to $\theta(\widehat{\mu})$ can be constructed as $\beta(\mu_0) = E\left\{\theta(\widehat{\mu}) - \theta(\mu_0)|\mu_0\right\}$ and a bias corrected estimator is thus obtained by correcting for the bias, $\theta(\widehat{\mu})_c = \theta(\widehat{\mu}) + \beta(\mu_0)$.

Since we observe only one realization $\mathbf{x} = \{x_1, ..., x_n\}$ from the unknown DGP μ_0, the sampling from the unknown probability distribution is replaced by resampling from the empirical distribution of the observed data.

Given a choice of $\widehat{F}$ a random bootstrap sample $\mathbf{X}^*$ of size n with realization $\mathbf{x}^* = (x_1^*, x_2^*, ..., x_n^*)$ is drawn with replacement from the original sample. [2]

[1] In what follows $E_\mu(.)$ stands for the expectation under the probability distribution F. This is used interchangeably with the notation $E(.|\mu)$.

[2] Here we abstract from the debate about the how much sampling of $\mathbf{x}$ from $\widehat{F}$. Without loss of generality, we set the bootstrap sample size equal to the original sample size.

The bootstrap principle approximates the sampling distribution of $R_{t(\mu_0)}(\widehat{\mu}, \mu_0)$ by the bootstrap distribution of $R_{t(\widehat{\mu})}(\widehat{\mu}^*, \widehat{mu})$. 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 $\widehat{F}^*(y) = \frac{1}{n}\sum_{i=1}^n I(X_i^* \leq y)$.

Let $\widehat{\mu}^*$ be the data generating process of the bootstrap sample and let $\theta(\widehat{\mu}^*)$ be the estimator $\theta(\mu)$ applied to the bootstrap sample $\mathbf{x}^* = \{x_1^*, ..., x_n^*\}$.

To fix ideas, define a function $H(x)$ which measures the amount of uncorrected bias in $\theta(\widehat{\mu})$ that remains after accounting for x,

$$H(x) \quad = \quad E\left\{\theta(\widehat{\mu}) - \theta(\mu_0) - x|\mu_0\right\}; \tag{3}$$

and let $\widehat{G}(x) = E\left\{\theta(\widehat{\mu}^*) - x|\widehat{\mu}\right\}$.

To estimate the bias of $\theta(\widehat{\mu})$, we would like to find the value t_0 such that

$$H(t_0) \quad = \quad E\left\{\theta(\widehat{\mu}) - \theta(\mu_0) + t_0|\mu_0\right\} = 0. \tag{4}$$

Efron's bootstrap method for calculating bias corrected estimators replaces the population equation (3) by the sample equation

$$\widehat{H}(x) \quad = \quad E\left\{\theta(\widehat{\mu}^*) - \theta(\widehat{\mu}) + x|\widehat{\mu}\right\}. \tag{5}$$

Definition 2.1. *The single bootstrap bias corrected estimator is defined as* $\theta(\widehat{\mu})_{bc} = \theta(\widehat{\mu}) + \widehat{t}^*$ *where* $\widehat{t}^* = \theta(\widehat{\mu}) - \widehat{G}^{-1}(0)$

Proof 1. *The theoretical bootstrap estimate for the bias in (4) is given by* x^*

$$H(t^*) \quad = \quad E\left\{\theta(\widehat{\mu}) - \theta(\mu_0) + t^*|\widehat{\mu}\right\} = 0 \tag{6}$$

However, since we observe only one realization $\mathbf{x} = (x_1, ..., x_n)$ *from the unknown DGP* μ_0*, the theoretical bias can be estimated by* $\widehat{t}^*$ *by replacing the population equation in (6) by the bootstrap equation*

$$\widehat{H}(\widehat{t}^*) \quad = \quad E\left\{\theta(\widehat{\mu}^*) - \theta(\widehat{\mu}) + \widehat{t}^*|\widehat{\mu}\right\} = 0 \tag{7}$$

An estimate of $\widehat{t}^*$ is calculated using Monte Carlo methods. Given the original sample, B bootstrap samples are generated from $\widehat{F}$. For each bootstrap sample j, $(j = 1, ..., B)$, a realized value $\widehat{\theta}^*$ of $\theta(\widehat{F})$. Let $\widehat{t}^*$ be the estimate for the bias in (8) then,

$$\widetilde{\widehat{t}}^* \quad = \quad \widehat{\theta} - \frac{1}{n}\sum_{j=1}^{n}\widehat{\theta}_j^* \tag{8}$$

The bootstrap bias-corrected estimator for $\theta(\mu_0)$ is thus,

$$\widehat{\theta}_{bc} \quad = \quad \theta(\widehat{\mu}) + \widetilde{\widehat{t}}^* = 2 \cdot \theta(\widehat{\mu}) - E_{\widehat{\mu}}\theta(\widehat{\mu}^*) \tag{9}$$

The amount of uncorrected bias in $\widehat{\theta}_{bc}$ is of order $O(n^{-2})$, that is

$$E_{\mu_0}\left\{\widehat{\theta}_{bc} - \theta(\mu_0)\right\} \quad = \quad O(n^{-2})$$

An improvement compared to the original estimator $\theta(\widehat{\mu})$

$$E_{F_0}\left\{\theta(\widehat{\mu})-\theta(\mu_0)\right\} \quad = \quad O(n^{-1})$$

For a thorough analysis of the bootstrap refinements see among others, Horowitz Horowitz [2001], Hall Hall and Horowitz [1996], Efron Efron [1987, 1979] and, Efron and Tibshirani Efron and Tibshirani [1986].

2.2. Double Bootstrap Bias Correction

Beran [1988, 1987] introduced the idea of repeated prepivoting by mapping a test statistic $\tau_{n,j}$ into the 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. Beran argued that the null distribution of $\tau_{n,j}$ is less strongly dependent of the parameters indexing the unknown probability distribution F_0. This idea is similar to Hall [1986] iterated bootstrap. In estimating the P value of a test statistic, he shows that the accuracy of the approximation using the r^{th} (iterated) bootstrap critical value is of order $O(n^{-(r+1)/2})$.

In this section, we use results from Shi [1992] to derive a double bootstrap equation for bias estimation. Shi showed that the double bootstrap principle can be used without the need of a pivot. He showed that the the double bootstrap one sided confidence interval has coverage probability $O(n^{-1})$ which faster than the percentile bootstrap. We borrow heavily from Shi's notation as well as that of Hall [1992].

Although Shi outlines the treatment for P value and confidence limits, the results in this section adapts and extends his double bootstrap equation to bias correction.

Let $\widehat{F}^{**}$ be the empirical distribution of the bootstrap sample $\mathbf{X}^{**}$ randomly drawn from the first level bootstrap distribution $\widehat{F}^{*}$ and let $\widehat{\mu}^{**}$ be the DGP determined by $\widehat{F}^{**}$.

In the following we follow Hall Hall [1986] and Shi Shi [1992] notation and methodology to describe the double bootstrap for estimating the bias.

Because the likelihood function of θ differs from the conditional density function of $\theta(\widehat{\mu}^{*})$, the bootstrap bias estimator $\widehat{t}^{*}$ in Lemma (2.1) although it satisfies (6) does not necessarily satisfy the population equation in (4),

$$E\left\{\theta(\widehat{\mu})-\theta(\mu_0)+\widehat{t}^{*}|\mu_0\right\} \quad \neq \quad 0 \tag{10}$$

The idea of the double bootstrap is to estimate the deviation from zero in (10) by using a second level bootstrap. In other terms, by iterating the bootstrap principle a second time, one is attempting to estimate the perturbation to $\widehat{t}^{*}$ needed to satisfy equation (4) with equality.

Definition 2.2. *Let $\widehat{H}^{*}(t)=E\left\{\theta(\widehat{\mu}^{**})-\theta(\widehat{\mu}^{*})+t|\widehat{\mu}^{*}\right\}$ and define $\beta^{*}=\widehat{H}^{*}(t^{*}_{\beta^{*}})$ where $t^{*}_{\beta^{*}}=\widehat{H}^{-1}(0)$; the double bootstrap bias corrected estimator is defined as*

$$\theta(\widehat{\mu})_{Dbc}=\theta(\widehat{\mu})+t^{**};$$

where $t^{**}=H^{-1}(\beta^{*})=\theta(\widehat{\mu})-\widehat{G}^{-1}(\beta^{*})$

Proof 2. *We want to estimate the deviation from zero of the population equation (10) and use it as a bias adjustment. Ideally, we want to find the bias adjustment also called the "calibrating coefficient" β such that,*

$$\begin{aligned} E\left\{\theta(\widehat{\mu}) - \theta(\mu_0) + t_\beta | \mu_0\right\} &= 0 \qquad (11)\\ E\left\{\theta(\widehat{\mu}^*) - \theta(\widehat{\mu}) + t_\beta | \widehat{\mu}\right\} &= \beta \qquad (12)\end{aligned}$$

In the first equation $\theta(\widehat{\mu})$ is random and depends upon μ_0 and the chosen calibrating coefficient β, while in the second equation $\theta(\widehat{\mu})$ is fixed and $\theta(\widehat{\mu}^*)$ is random and depends upon the DGP $\widehat{\mu}^*$.

The purpose is to solve the system of equations for the value of β. To do so, the bootstrap principle is used to provide an estimate β^* that satisfies,

$$\begin{aligned} E\left\{\theta(\widehat{\mu}^*) - \theta(\widehat{\mu}) + t^*_{\beta^*} | \widehat{\mu}\right\} &= 0 \qquad (13)\\ E\left\{\theta(\widehat{\mu}^{**}) - \theta(\widehat{\mu}^*) + t^*_{\beta^*} | \widehat{\mu}^*\right\} &= \beta^* \qquad (14)\end{aligned}$$

We can rewrite equation (14) as $t^*_{\beta^*} = \widehat{H}^{*-1}(\beta^*) = \theta(\widehat{\mu}^*) - \widehat{G}^{*-1}(\beta^*)$. Solving for β^* in (13) gives,

$$E\left\{\theta(\widehat{\mu}^*) - \theta(\widehat{\mu}) + \theta(\widehat{\mu}^*) - \widehat{G}^{*-1}(\beta^*) | \widehat{\mu}\right\} = 0 \qquad (15)$$

or equivalently, $\widehat{H}\left(\theta(\widehat{\mu}^*) - \widehat{G}^{*-1}(\beta^*)\right) = 0$. Finally,

$$\beta^* = \widehat{G}^*\left(2 \cdot \theta(\widehat{\mu}^*) - \theta(\widehat{\mu})\right) \qquad (16)$$

To find the double bootstrap estimate for the bias, we then solve for t^{**} such that,

$$\begin{aligned} \widehat{H}(t^{**}) &= \beta^*\\ t^{**} &= \widehat{H}^{-1}(\beta^*) = \theta(\widehat{\mu}) - \widehat{G}^{-1}(\beta^*)\end{aligned}$$

Noting that $\widehat{G}^{-1}(\beta^*) = E(\theta(\widehat{\mu}^*) - \beta^* | \widehat{\mu})$ and $\beta^* = E\left\{\theta(\widehat{\mu}^{**}) - 2\theta(\widehat{\mu}^*) + \theta(\widehat{\mu}) | \widehat{\mu}^*\right\}$, we can finally rewrite the double bootstrap bias correction as,

$$t^{**} = 2 \cdot \theta(\widehat{\mu}) - 3 \cdot E\left\{\theta(\widehat{\mu}^*) | \widehat{\mu}\right\} + E\left\{\theta(\widehat{\mu}^{**}) | \widehat{\mu}\right\}$$

The expression for the double bootstrap bias corrected estimator for $\theta(\mu_0)$ is therefore

$$\theta(\widehat{\mu})_{Dbc} = 3 \cdot \theta(\widehat{\mu}) - 3 \cdot E\left\{\theta(\widehat{\mu}^*) | \widehat{\mu}\right\} + E\left\{\theta(\widehat{\mu}^{**}) | \widehat{\mu}\right\} \qquad (17)$$

In fact, Hall Hall [1992] showed that the r^{th} iterated bootstrap bias corrected estimator can be written as,

$$\theta(\widehat{\mu})_{rbc} = \sum_{j=1}^{r+1} \begin{pmatrix} r+1 \\ j \end{pmatrix} E\left\{\theta(\widehat{\mu}^{r \cdot *}) | \widehat{\mu}\right\} \qquad (18)$$

where $\widehat{\mu}^{r \cdot *}$ is the DGP for the r^{th} bootstrap resample.

It is worth noting that the proof of Lemma 2.2 follows the treatment of confidence limits in Shi Shi [1992]. Although it differs from Hall's exposition of additive bias correction, the two basically lead to the exact same bias correction. The proof of the equivalence of the two is deferred to the appendix.

2.3. Implementing the double bootstrap bias correction

If the first level bootstrap is accurate in estimating the distribution of the random variable $\theta(\widehat{\mu})$ then the bias adjustment β^* will be very small. In reality however, the accuracy of the bias correction increases with the double bootstrap. Indeed Hall [1986] argues that for the r^{th} level bootstrap ($r = 2$ corresponds to the double bootstrap) the increase of the accuracy of the iterated bootstrap is of order $O(n^{-r})$. To implement Hall's double bootstrap bias correction, the Monte Carlo algorithm involves bootstrapping within the first level bootstrap.

1. For each (first level) bootstrap replication j ($j = 1, .., B_1$), a bootstrap sample $\mathbf{x}^*$ is drawn using the empirical distribution $\widehat{F}$. Calculate the bootstrap realized value $\widehat{\theta}^*_j$ of the statistic $\theta(\widehat{F}^*)$.
   The simulated data $\mathbf{x}^*$ is then used to construct a second level bootstrap data generating process with CDF $\widehat{F}^*$.
2. For each (second level) bootstrap replication l ($l = 1, .., B_2$), a bootstrap sample $\mathbf{x}^{**}$ is drawn from $\mathbf{x}^*$ using $\widehat{F}^*$. Calculate the bootstrap realized value $\widehat{\theta}^{**}_{j,l}$ of the statistic $\theta(\widehat{F}^{**})$. An estimate for the second level bias correction $\widehat{b}^{**}$ is

$$\widehat{b}^{**}_j \quad = \quad \widehat{\theta}^*_j - \frac{1}{B_2}\sum_{l=1}^{B_2}\widehat{\theta}^{**}_{j,l}. \tag{19}$$

3. After all bootstrapping operations are complete, we can calculate an estimate for the first level bootstrap bias,

$$\widehat{b}^* \quad = \quad \widehat{\theta} - \frac{1}{B_1}\sum_{j=1}^{B_1}\widehat{\theta}^*_j; \tag{20}$$

   and an estimate for the bias adjustment $\widetilde{b}$,

$$\widehat{\widetilde{b}} \quad = \quad \widehat{\theta} - \frac{1}{B_1}\sum_{j=1}^{B_1}\widehat{\theta}^*_j - \frac{1}{B_1}\sum_{j=1}^{B_1}\widehat{b}^{**}_j = \widehat{b}^* - \frac{1}{B_1}\sum_{j=1}^{B_1}\widehat{b}^{**}_j. \tag{21}$$

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 $H(.)$. Depending on the model and the estimation method indeed this can be computationally cumbersome.

3. Fast Double Bootstrap

3.1. Higher order refinements of the FDB

Davidson and MacKinnon [2007, 2002b] developed a new procedure for approximating the bootstrap P values and rejection probabilities they called *fast double bootstrap (FDB).* These approximations are more accurate than the single bootstrap estimates but are less computationally demanding than the double bootstrap. The procedure proposed in this study follows the same argument as in Davidson and MacKinnon [2007, 2002b]. Instead of

Definition 3.1. *Let $b_{\theta(\widehat{\mu})}$ be the bias associated with $\theta(\widehat{\mu})$ such that $\widehat{H}[b_{\theta(\widehat{\mu})}] = 0$. Let $x^*_{b(\theta(\widehat{\mu}))} = \theta(\widehat{\mu}^*) - \widehat{H}^{*-1}[b_{\theta(\widehat{\mu})}]$, then the fast double bootstrap bias corrected estimator for $\theta(\mu)$ is defined as,*

$$\theta(\widehat{\mu})_{Fbc} = \theta(\widehat{\mu}) + b^* + b_{\theta(\widehat{\mu})};$$

where $b^* = \widehat{H}[\theta(\widehat{\mu}) - x^*(b_{\theta(\widehat{\mu})})]$

In this section we describe a technique to increase the accuracy of bootstrap bias estimation in the same spirit of the *FDB* of Davidson and MacKinnon.

3.2. Implementing the fast double bootstrap bias correction

1. For each bootstrap replication j ($j = 1, .., B_1$), a bootstrap sample $\mathbf{x}^*$ is drawn using the empirical distribution $\widehat{F}$. Calculate the bootstrap realized value $\widehat{\theta}^*_j$ of the statistic $\theta(\widehat{F}^*)$.
   The simulated data $\mathbf{x}^*$ is then used to construct a second level bootstrap data generating process with CDF $\widehat{F}^*$. One bootstrap sample $\mathbf{x}^{**}$ is drawn from $\mathbf{x}^*$ using $\widehat{F}^*$. Calculate the bootstrap realized value $\widehat{\theta}^{**}_j$ of the statistic $\theta(\widehat{F}^{**})$.
2. After all bootstrapping operations are complete, we have two series of iterates, $\widehat{\theta}^*_j$ and $\widehat{\theta}^{**}_j$. The first level bootstrap bias is calculated in similar way as in (20).
3. Instead of using the mean of the distribution of $\widehat{\theta}^*_j$ to calculate second level bias, we compute a value Q^{**} such as

$$x^{**} = \frac{1}{B_1}\sum_{j=1}^{B_1}\widehat{\theta}^{**}_j + \widehat{b}^*, \tag{22}$$

   and compute and estimate for $\widetilde{\eta}$,

$$\widehat{\widetilde{\eta}} = x^{**} - \frac{1}{B_1}\sum_{j=1}^{B_1}\widehat{\theta}^*_j \tag{23}$$

As in Davidson and MacKinnon [2007, 2002b] , this algorithm requires only $2B_1$ visits to the statistic of interest, in this case the estimator for $\theta(F)$.

4. The Monte Carlo Environment

4.1. Linear IV model

Consider the simple linear IV model of Hahn and Hausman [2002] and as presented in Guggenberger [2008]

$$y_i = \theta x_i + \epsilon_i \tag{24}$$
$$x_i = z_i\pi + v_i \qquad i = 1, ..., n \tag{25}$$
$$\tag{26}$$

For simplicity, we assume that the endogenous variable in the left-hand side of (24) is a scalar. The scalar regressor x_i is endogenous and accepts the reduced form in (25). 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, $\mathbb{R}^2 = \frac{\pi'\pi}{1+\pi'\pi}$.

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 $\mathbb{R}^2$ is thus related to the relevance of the instruments and to the number of instruments in the simple equation ,

$$\mathbb{R}^2 = \frac{K.\eta^2}{1+K.\eta^2} \tag{27}$$

The IV estimator is defined as

$$\widehat{\theta}_{IV} = (\mathbf{x}'P_z\mathbf{x})^{-1}\mathbf{x}'P_z\mathbf{y} \tag{28}$$

where P_z is the linear projection matrix defined by $P_z = \mathbf{z}(\mathbf{z}'\mathbf{z})^{-1}\mathbf{z}'$. This estimator is consistent and asymptotically normally distributed. The finite sample bias is however dependent on ρ, π and K. Indeed, Rothenberg [1983] provides an expression for the approximate bias which clearly shows how these parameters come into play:

$$bias(\widehat{\theta}_{IV}) = \frac{(K-2).\rho}{n.(\pi'\mathbf{z}'\mathbf{z}\pi)^{-1}}. \tag{29}$$

The data are simulated to represent cases of weak instruments (low $\mathbb{R}^2$), and cases of severe endogeneity (high ρ). The degree of overidentification (number of instruments) does also play a role in the tradeoff between bias and efficiency for IV estimation. We simulate data for the possible parameter combinations of $n \in \{100, 200\}$, $K \in \{1, 5, 20\}$, $\rho \in \{0.3, 0.9\}$ and $\mathbb{R}^2 \in \{0.001, 0.1\}$. In all the experiments, θ is set to zero. One important result in the IV estimation is that the m^{th} moment exists if and only if $m < K$. In the context of our simulation experiments, for $K = 1$ the distribution of IV estimator for β does not have any moments. In this case the results on the mean and standard errors need to be interpreted cautiously. For these reasons, we will be reporting the Median bias for analysis.

4.2. Nonlinear GMM

We investigate bias estimation properties of the fast method proposed in this paper in the GMM estimation of nonlinear conditional moment model. As an example we study the C-CAPM of Hansen [1982].

Table 1: $\mathbb{R}^2 = 0.10$, $n = 100$

	Panel A:$\rho = 0.9$				Panel B:$\rho = 0.3$			
	Mean	Median	Std	RMSE	Mean	Median	Std	RMSE
		$K = 1$				$K = 1$		
$\widehat{\beta}$	-0.0095	0.0005	1.34	1.34	-0.0067	0.0254	0.4161	0.4162
$\widehat{\beta}_{bc}$	-0.1625	0.0579	5.400	5.40	-0.0365	0.0665	3.3422	3.342
$\widehat{\beta}_{Dbc}$	-0.3147	0.0989	11.03	11.03	-0.0668	0.1042	6.5972	6.597
$\widehat{\beta}_{Fbc}$	-0.5436	0.0971	18.52	18.53	-0.2727	0.0910	11.94	11.95
		$K = 5$				$K = 5$		
$\widehat{\beta}$	0.0926	0.0994	0.1074	0.1418	0.0815	0.0924	0.2847	0.2961
$\widehat{\beta}_{bc}$	0.0511	0.0682	0.1593	0.1673	0.0364	0.0586	0.3879	0.3896
$\widehat{\beta}_{Dbc}$	0.0097	0.0366	0.2140	0.2142	-0.0086	0.0240	0.4978	0.4979
$\widehat{\beta}_{Fbc}$	0.0294	0.0544	0.1973	0.1995	0.0190	0.0488	0.4548	0.4552
		$K = 20$				$K = 20$		
$\widehat{\beta}$	0.1696	0.1719	0.0568	0.1789	0.1693	0.1692	0.1710	0.2406
$\widehat{\beta}_{bc}$	0.1513	0.1583	0.0821	0.1722	0.1385	0.1400	0.2355	0.2732
$\widehat{\beta}_{Dbc}$	0.1331	0.1426	0.1083	0.1716	0.1078	0.1107	0.3036	0.3222
$\widehat{\beta}_{Fbc}$	0.1404	0.1506	0.0990	0.1718	0.1214	0.1258	0.2764	0.3019

The experimental design will borrow heavily from the extensive body of empirical work investigating inference in the consumption asset pricing model using GMM methods.

The Euler equation in the C-CAPM for an economy with m assets and assuming that the preferences are of the constant relative risk aversion type is

$$E_t \left\{ \alpha(c_{t+1})^{-\gamma} \otimes R_{t+1} - \iota_m \right\} = \mathbf{0} \tag{30}$$

where c_t is the growth rate of consumption, R_t is the m dimensional vector of gross stock returns, γ is the risk aversion parameter (> 0), α is the impatience parameter, and ι_m is an $m \times 1$ vector of ones. The one period gross return from holding one unit of stock j is defined as:

$$R_{j,t+1} = \frac{P_{j,t+1} + D_{j,t+1}}{P_{j,t}}$$

where $D_{j,t+1}$ is the dividend yield on stock j from period t to $t+1$.

Given a set z_t of q instruments, available at time t, a family of population orthogonality conditions can be constructed based on the following moments functions[3]:

$$\begin{aligned} g_t(\theta) &= \left(\alpha(c_{t+1})^{-\gamma} \otimes R_{t+1} - \iota_m\right) \otimes z_t \\ g(\theta) &= \frac{1}{n} \sum_{t=1}^{n} g_t(\theta) \end{aligned}$$

The continuously updated GMM of Hansen *et. al* Hansen et al. [1996] is defined by,

$$\widehat{\theta} = \arg\min_{\theta} g(\theta)' \left[\frac{1}{n} \sum_{t=1}^{n} g_t(\theta) g_t(\theta)' \right]^{-1} g(\theta) \tag{31}$$

[3] $g_t(\theta_0) = h(x_{t+1}, \theta_0) \otimes z_t$ stands for the Kronecker product of the $m \times 1$ vector $h(x_{t+1}, \theta_0)$ and the $q \times 1$ vector of instruments z_t. The product is an $mq \times 1$ vector.

Table 2: $\mathbb{R}^2 = 0.001$, $n = 100$

	Panel A:$\rho = 0.9$,				Panel B:$\rho = 0.3$			
	Mean	Median	Std	RMSE	Mean	Median	Std	RMSE
	$K = 1$				$K = 1$			
$\widehat{\beta}$	0.2706	0.1837	17.51	17.51	1.1204	0.2753	15.92	15.96
$\widehat{\beta}_{bc}$	0.2901	0.1798	35.23	35.23	2.2363	0.3581	38.30	38.36
$\widehat{\beta}_{Dbc}$	0.3122	0.1992	53.19	53.19	3.3458	0.4183	63.39	63.48
$\widehat{\beta}_{Fbc}$	-0.4274	0.1584	58.83	58.83	2.5317	0.3965	84.19	84.22
	$K = 5$				$K = 5$			
$\widehat{\beta}$	0.2203	0.2189	0.1345	0.2582	0.2663	0.2452	0.51	0.57
$\widehat{\beta}_{bc}$	0.2237	0.2216	0.2207	0.3142	0.2661	0.2537	0.83	0.87
$\widehat{\beta}_{Dbc}$	0.2270	0.2210	0.3087	0.3832	0.2659	0.2565	1.16	1.19
$\widehat{\beta}_{Fbc}$	0.2267	0.2191	0.2926	0.3701	0.2666	0.2558	1.09	1.12
	$K = 20$				$K = 20$			
$\widehat{\beta}$	0.2194	0.2184	0.0547	0.2262	0.2587	0.2556	0.1996	0.3268
$\widehat{\beta}_{bc}$	0.2205	0.2203	0.0822	0.2353	0.2571	0.2626	0.3009	0.3958
$\widehat{\beta}_{Dbc}$	0.2215	0.2215	0.1108	0.2477	0.2554	0.2651	0.4061	0.4797
$\widehat{\beta}_{Fbc}$	0.2212	0.2208	0.1019	0.2436	0.2559	0.2608	0.3735	0.4528

To simulate series of consumption growth and stock returns which satisfy the set of moments conditions in (30), we follow the approach of **?**. This approach assumes that the state variables (consumption growth c_t, and dividend growths, $X_{j,t} = D_{j,t}/D_{j,t-1}$) are jointly stationary first order Markov processes. An $9 - state$ Markov chain is fitted to the consumption and dividends growth so as to approximate the first order vector autoregression (VAR)

$$\mathbf{X_t} = \mu + \mathbf{\Phi X_{t-1}} + \epsilon_\mathbf{t}$$

where $\mathbf{X_t} = (log(c_t), log(X_{1,t}), ..., log(X_{m,t}))'$ and $\epsilon_\mathbf{t}$ are independent and identically distributed with $E_{t-1}(\epsilon_\mathbf{t}) = 0$ and $V(\epsilon_\mathbf{t}) = \Omega$. Our economy is similar to the one described by Kocherlakota [1990] with complete and frictionless markets with three assets: the Risk free R^f which pays one unit of consumption, the market portfolio MP which pays C_t units of consumption in period t and the stock market SM with dividend payoffs D_t in period t.

Wright [2003] addressed issues of identification in the C-CAPM and proposed a test for detecting lack of identification. We use his Monte Carlo specifications in our problem to detect the effect of identification on the ability of the bootstrap to approximate the estimation bias. The models studied by Wright [2003] are presented in Table 5. Model (FR) represent the case of fully identified model, NRF stands for the case where the parameter γ is weakly identified and finally RF represent a situation where the model is not fully identified.

5. Monte Carlo Results

The results reported in this section are based on 1,000 simulation repetitions. There are a number of studies which explored the choice of the number of bootstraps and how it

Table 3: $\mathbb{R}^2 = 0.1$, $n = 200$

	Panel A:$\rho = 0.9$,				Panel B:$\rho = 0.3$			
	Mean	Median	Std	RMSE	Mean	Median	Std	RMSE
	$K = 1$				$K = 1$			
$\widehat{\beta}$	-0.0200	0.0023	0.1207	0.1223	-0.0067	0.0254	0.4161	0.4162
$\widehat{\beta}_{bc}$	0.0018	0.0350	0.3456	0.3456	-0.0365	0.0665	3.3422	3.3424
$\widehat{\beta}_{Dbc}$	0.0241	0.0613	0.6498	0.6503	-0.0668	0.1042	6.5972	6.5976
$\widehat{\beta}_{Fbc}$	-0.0151	0.0469	1.1860	1.1861	-0.2727	0.0910	11.94	11.95
	$K = 5$				$K = 5$			
$\widehat{\beta}$	0.0509	0.0603	0.0800	0.0948	0.1390	0.1472	0.1511	0.2053
$\widehat{\beta}_{bc}$	0.0147	0.0331	0.1120	0.1129	0.0968	0.1097	0.2050	0.2267
$\widehat{\beta}_{Dbc}$	-0.0213	0.0062	0.1465	0.1481	0.0545	0.0686	0.2606	0.2662
$\widehat{\beta}_{Fbc}$	0.0021	0.0267	0.1321	0.1321	0.0743	0.0866	0.2361	0.2476
	$K = 20$				$K = 20$			
$\widehat{\beta}$	0.1241	0.1253	0.0455	0.1322	0.1264	0.1249	0.1459	0.1930
$\widehat{\beta}_{bc}$	0.0920	0.0942	0.0633	0.1117	0.0858	0.0858	0.1946	0.2127
$\widehat{\beta}_{Dbc}$	0.0600	0.0646	0.0820	0.1017	0.0452	0.0500	0.2452	0.2493
$\widehat{\beta}_{Fbc}$	0.0748	0.0792	0.0745	0.1056	0.0656	0.0659	0.2212	0.2307

affects the theoretical predictions abut the bootstrap approximation. Most of these studies considered the case of bootstrap tests and confidence intervals. Davidson and MacKinnon [2000] argue that the outcome of the bootstrap test will depend on the sequence of random numbers used to generate the bootstrap samples, and it necessarily results in some loss of power. They propose a data dependent pretest procedure for choosing the number of bootstrap samples so as to minimize power loss. This procedure also depends on the the nominal level π.

The choice of the number of inner bootstraps is equally important. Lee and Young [2002] studied the effect of experimental randomness on coverage error of double bootstrap confidence intervals. They show that to ensure that the coverage error in the Monte Carlo remains of the same order as that of the theoretical one, B_1 and B_2 must be of larger order than the sample size n, of order n^4 and n^2 respectively in the two-sided case and of order n^2 and n respectively in the one-sided case.

Our choice of B_1 and B_2 is not justified by any data dependent measure of optimality. Because of the computational cost involved with increased number of bootstraps, we restrict our selves to the choice of $B_1 = 499$ and $B_2 = 399$ (considered as reasonable in Davidson and MacKinnon [2007]).

For both the linear IV and nonlinear GMM, we calculate the mean bias, median bias, standard errors (*std*)and root mean square error (*RTMSE*) of the estimator $\widehat{\theta}$, the single bias corrected (*SBBC*) estimator $\widehat{\theta}_{bc}$, the double bootstrap bias-corrected (*DBBC*) estimator $\widehat{\theta}_{Dbc}$ (Lemma 2.2) and the proposed fast bootstrap bias-corrected (*FBBC*) estimator $\widehat{\theta}_{Fbc}$ (Lemma 3.1)

5.1. Linear IV

In this section, we discuss the performance of the proposed bootstrap bias-correction relative to the single and double bootstrap corrected estimators. We also discuss the effect of n, $\mathbb{R}^2$ and K on the performance of each of the estimators.

Table 4: $\mathbb{R}^2 = 0.001$, $n = 200$

	Panel A:$\rho = 0.9$,				Panel B:$\rho = 0.3$			
	Mean	Median	Std	RMSE	Mean	Median	Std	RMSE
		$K = 1$				$K = 1$		
$\widehat{\beta}$	-0.4543	0.2047	27.74	27.74	2.810	0.2123	44.76	44.85
$\widehat{\beta}_{bc}$	-1.0148	0.2066	55.76	55.77	6.993	0.2090	102.13	102.37
$\widehat{\beta}_{Dbc}$	-1.5765	0.2014	83.98	83.99	11.18	0.1475	166.54	166.91
$\widehat{\beta}_{Fbc}$	-1.3936	0.1888	85.24	85.25	12.66	0.1997	200.89	201.29
		$K = 5$				$K = 5$		
$\widehat{\beta}$	0.2146	0.2179	0.1515	0.2627	0.2471	0.2781	0.4819	0.5415
$\widehat{\beta}_{bc}$	0.2137	0.2197	0.2587	0.3356	0.2463	0.2801	0.7790	0.8170
$\widehat{\beta}_{Dbc}$	0.2129	0.2187	0.3679	0.4251	0.2456	0.2872	1.0821	1.1094
$\widehat{\beta}_{Fbc}$	0.2132	0.2199	0.3534	0.4128	0.2486	0.2839	**1.0199**	**1.0498**
		$K = 20$				$K = 20$		
$\widehat{\beta}$	0.2181	0.2176	0.0590	0.2259	0.2727	0.2617	0.1775	0.3254
$\widehat{\beta}_{bc}$	0.2191	0.2185	0.0901	0.2369	0.2765	0.2641	0.2643	0.3825
$\widehat{\beta}_{Dbc}$	0.2202	0.2207	0.1217	0.2516	0.2804	0.2663	0.3533	0.4511
$\widehat{\beta}_{Fbc}$	0.2199	0.2197	0.1123	0.2469	0.2793	0.2651	**0.3248**	**0.4284**

Tables 1-4 contain the simulation results for all the combinations of $n \in \{100, 200\}$, $\mathbb{R}^2 \in \{0.001, 0.1\}$, $K \in \{1, 5, 10\}$ and $\rho \in \{0.3, 0.9\}$. The columns of each table report the sample mean, median,[4] *std* and *RTMSE* of $\widehat{\beta}$, $\widehat{\beta}_{bc}$, $\widehat{\beta}_{Dbc}$ and $\widehat{\beta}_{Fbc}$.

The results for the linear IV estimator $\widehat{\beta}$ are conventional. As the degree of overidentification increases (higher K), there is an increase in the mean and median bias while the standard errors go down. The bias and *RTMSE* are higher for models with high endogeneity (high ρ in Panel A versus low ρ in Panel B). The bias is especially pronounced for As expected from the theoretical predictions in the existing literature about the iterated bootstrap, the mean bias of the double bootstrap is smaller than that of the single bootstrap. However, this gain in bias does not offset the higher levels of standard errors and therefore increased *RTMSE*.

Phillips [1980] derived the exact probability distribution function for the IV estimator in simultaneous equations models. The leading term in the density is proportional to a multivariate t-distribution and reveals that the integer moments of the IV estimator exist up to the degree of over-identification. For the just identified case ($K = 1$), the first and second moments of the IV estimator do not exist. This suggest additional finite sample problems with the just identified model.

Results in Table 1 show that the proposed fast bias corrected estimator $\widehat{\beta}_{Fbc}$ outperforms the single bootstrap bias corrected $\widehat{\beta}$ in terms of mean and median bias. The double bootstrap estimator $\widehat{\beta}_{Dbc}$ further shrinks this bias outperforming the proposed estimator. Our estimator although less precise than $\widehat{\beta}$ outperforms $\widehat{\beta}_{Dbc}$ in terms of *std* and *RTMSE*.

The results are even more promising for the *FBBC* estimator when the sample size is increased from $n = 100$ to $n = 200$. Table 3 shows that for models where instruments'

[4] Because the true value of β in the DGP is zero, the mean and median are also a measure of the mean and median bias.

Table 5: Parameter Settings for the monte Carlo experiments: $\gamma = 1.30$, $\alpha = .97$.

Model	Φ	Ω	μ
FR	$\begin{pmatrix} -.50 & 0 \\ 0 & -.50 \end{pmatrix}$	$\begin{pmatrix} .01 & .00 \\ .00 & .01 \end{pmatrix}$	$\begin{pmatrix} .00 \\ .00 \end{pmatrix}$
NRF	$\begin{pmatrix} -.161 & .017 \\ .414 & .117 \end{pmatrix}$	$\begin{pmatrix} .00120 & .00177 \\ .00177 & .01400 \end{pmatrix}$	$\begin{pmatrix} 0.021 \\ 0.004 \end{pmatrix}$
RF	$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$	$\begin{pmatrix} .00120 & .00177 \\ .00177 & .0146 \end{pmatrix}$	$\begin{pmatrix} 0.018 \\ 0.013 \end{pmatrix}$

Note: The VAR model is $\mathbf{X_t} = \mu + \mathbf{\Phi X_{t-1}} + \epsilon_\mathbf{t}$, where $\mathbf{X_t} = (log(c_t), log(x_t))'$ and $V(\epsilon_\mathbf{t}) = \Omega$. The economy is calibrated with coefficient of relative risk aversion γ, and discount factor β. $M1$ and $M2$ are benchmark models where the serial autocorrelation in the VAR is weak and the dividinds and consumption growth series are uncorrelated. $M3$ and $M5$ are considered by Kocherlakota [1990],Hansen et al. [1996] and Wright [2003] . also considers models Models $M4$ and $M6$ introduce relatively strong serial correlation in the dividends series. These models are discusses further in the text.

Table 6: Monte Carlo results for γ

		Panel A: $n = 100$			Panel B: $n = 200$		
Model		$\widehat{\gamma}$	$\widehat{\gamma}_{bc}$	$\widehat{\gamma}_{Fbc}$	$\widehat{\gamma}$	$\widehat{\gamma}_{bc}$	$\widehat{\gamma}_{Fbc}$
FR	Mean	1.8568	1.1706	0.8509	1.5680	1.2013	1.0372
	Median	1.3918	1.0803	0.9531	1.3376	1.1782	1.2706
	Std	2.2604	3.2156	4.4741	1.2723	2.0983	3.2560
	RMSE	2.9252	3.4220	4.5543	2.0192	2.4179	3.4172
RF	Mean	0.0938	-0.0221	-0.1518	0.0794	-0.0713	-0.1794
	Median	0.0018	0.0243	-0.0135	-0.0011	-0.0867	-0.0968
	Std	1.7980	2.8409	4.1745	0.7828	1.4191	2.1654
	RMSE	1.8005	2.8409	4.1773	0.7869	1.4209	2.1728
NRF	Mean	6.6182	6.9244	6.5954	7.6538	8.3402	8.8194
	Median	7.6569	8.7511	8.5980	7.8630	8.3627	8.4250
	Std	4.3968	8.0732	12.3990	2.7785	4.8951	7.5716
	RMSE	0.1691	0.2152	14.0440	0.1691	0.2152	14.0440

relevance is not an issue, our proposed bias corrected estimator not only outperforms the single bootstrap in terms of mean bias but also surpasses the *DBBC*. *FBBC* outperformed the double bootstrap in terms of mean bias for the model with the highest degree of endogeneity (Panel A, case of $K = 5$). For the model with the largest number of (relevant) instruments, the precision of *FBBC*is even higher than that of the *SBBC* estimator (Panel A, case of $K = 20$).

For models with weak instrument problems, the main lesson from the simulation results is that the bootstrap does not help in reducing the finite sample bias. Tables 2-4 suggest that iterating the bootstrap *DBBC* is worse than the *SBBC* , which in terns does not improve upon the uncorrected estimator $\widehat{\beta}$. The proposed *FBBC* does not achieve any bias reduction but also does not do any worse than the *DBBC*. Increasing the sample size does not offset the lack of relevance of the instruments. However, the simulation results suggest that weak instruments are less of a problem when endogeneity is more pronounced.

For the just identified model, the results suggest that iterating the bootstrap principle

Table 7: Monte Carlo results for α

Model		Panel A: $n = 100$ $\widehat{\alpha}$	$\widehat{\alpha}_{bc}$	$\widehat{\alpha}_{Fbc}$	Panel B: $n = 200$ $\widehat{\alpha}$	$\widehat{\alpha}_{bc}$	$\widehat{\alpha}_{Fbc}$
FR	Mean	0.0237	0.0117	0.0053	0.0158	0.0085	0.0068
	Median	0.0136	0.0051	0.0062	0.0120	0.0087	0.0104
	Std	0.0370	0.0588	0.0910	0.0169	0.0339	0.0599
	RMSE	0.0440	0.0599	0.0911	0.0231	0.0349	0.0603
RF	Mean	0.0035	0.0005	0.0003	0.0018	-0.0013	-0.0032
	Median	0.0000	-0.0001	0.0003	-0.0000	-0.0016	-0.0017
	Std	0.0277	0.0465	0.0714	0.0142	0.0257	0.0400
	RMSE	0.0279	0.0465	0.0714	0.0143	0.0257	0.0401
NRF	Mean	0.1442	0.1491	0.1417	0.1634	0.1754	0.1838
	Median	0.1587	0.1772	0.1812	0.1658	0.1743	0.1725
	Std	0.0883	0.1552	0.2384	0.0555	0.0931	0.1411
	RMSE	0.1691	0.2152	0.2773	0.1725	0.1986	0.2317

is not a good idea. The mean and median bias of the *DBBC* estimator is higher than that of the *SBBC*. The same is true for the *RTMSE* of the two estimators. The proposed *FDBC* estimator outperforms the *DBBC* in terms of median bias but still worse than the *SBBC*.

5.2. Non-linear GMM

6. Conclusion

This paper has presented a new computationally efficient technique for bias correction which can be used instead of the computationally intensive double bootstrap. The theory predicts that iterating the bootstrap principle increases the accuracy of the bootstrap. This increased accuracy comes at an enormous computational cost. With B_1 first level bootstraps and B_2 second level bootstraps, the double bootstrap bias correction requires the computation of $B_1B_2 + B_1 + 1$ statistic. Our proposed estimator improves on the single bootstrap and is more precise than the double bootstrap and requires computing only $2B_1 + 1$ statistic.

We compare the relative performance of the single bootstrap (*SBBC*), the double bootstrap (*DBBC*) and the new proposed fast bootstrap (*FBBC*) bias-corrected estimators in terms of mean and median bias, standard errors and root mean square errors.

The *DBBC* has the lowest mean bias compared to *SBBC* and *FBBC* at a high cost of low precision due to increase noise introduced in the second level resampling. Our proposed *FBBC* has a mean bias slightly higher than *DBBC* but significantly lower than *SBBC*. Furthermore, the *FBBC* estimator is significantly more precise than the double bootstrap.

We also find that the performance of all the bootstrap estimators is not affected by the degree of endogeneity across the models with relevant instruments.

In terms of the ability of the bootstrap bias correction to remedy the problem of weak instruments, the finding for the linear IV show that the bootstrap is not successful in recentering the density of the estimator towards the true value. We find that in the case of weak identification, models with high degree of endogeneity have lower level of mean and median bias. This result warrants further investigation.

References

Beran, R., 1987. Prepivoting to Reduce Level Error in Confidence Sets. Biometrika 74, 457–468.

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

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

Davidson, R., MacKinnon, J. G., 2000. Bootstrap Tests: How Many Bootstraps? Econometric Reviews 19 (1), 55–68.

Davidson, R., MacKinnon, J. G., 2002a. Bootstrap j tests of Nonnested Linear Regression Models. Journal of Econometrics 109, 167–193.

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

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

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

Efron, B., 1987. Better Bootstrap Confidence Intervals. Journal od American Statistical Association 82, 171–200.

Efron, B., Tibshirani, R., 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.

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

Hahn, J., Hausman, J., 2002. A New Specification Test for the Validity of Instrumental Variables. Econometrica 70, 163–189.

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

Hall, P., 1992. The Bootstrap and Edgeworth Expansion. Springer-Verlag, New York.

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

Hall, P., Martin, M. A., 1988. On Bootstrap Resampling and Iteration. Biometrika 75, 661–671.

Hansen, L., Heaton, J., Yaron, A., 1996. Finite Sample Properties of Some Alternative gmm Estimators. Journal of Business and Economic Statistics 14 (3), 262–280.

Hansen, L. P., 1982. Large Sample Properties of Generalized Method of Moments Estimators. Econometrica 50 (4), 1029–1054.

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

Kocherlakota, N., 1990. On Tests of Representative Consumer Asset Pricing Models. Journal of Monetary Economics 26, 285–304.

Lee, S. M. S., Young, G. A., 2002. 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.

Phillips, P. C. B., 1980.

Rothenberg, T. J., 1983. Asymptotic Properties of some Estimators in Structural Models. In Studies in Econometrics Time Series and Multivariate Statistics Karlin S, Amemeya T, Goodman (eds) (Academic Press: New York).

Shi, S. G., 1992. Accurate and Efiicient Double-bootstrap Confidence Limit Method. Computational Statistics & Data Analysis 13, 21–32.

Wright, H., J., 2003. Detecting Lack of Identification in gmm. Econometric Theory 19, 322–330.

Appendix: Equivalence of Hall and Shi double-bootstrap

In what follows, we restate the results from Hall Hall [1992] for bias correction and compare the double bootstrap adjustment from both methods.

Hall's argument starts from the population equation

$$E_{\mu_0}\left\{\theta(\widehat{\mu}) - \theta(\mu_0) + T(\mu_0)\right\} \;=\; 0 \tag{32}$$

The solution $\widehat{t}^*$ to the sample equation $E_{\widehat{\mu}}\left\{\theta(\widehat{\mu}^*) - \theta(\widehat{\mu}) + t(\widehat{\mu})\right\} = 0$ does not necessarily satisfy (32) with equality. To improve the approximation, Hall introduces an additive[5]

[5] The perturbation can also be multiplicative.

perturbation to $\widehat{t}^*$ in the form, $U(\mu_0, \widehat{t}_0) = \widehat{t}^* + T(\mu_0)$ such that the population equation holds,

$$E_{\mu_0}\left\{\theta(\widehat{\mu}) - \theta(\mu_0) + U(\mu_0, \widehat{t}^*)\right\} = 0 \tag{33}$$

The sample equation corresponding to (33),

$$E_{\widehat{\mu}}\left\{\theta(\widehat{\mu}^*) - \theta(\widehat{\mu}) + \widehat{t}^{**} + \widetilde{t}\right\} = 0 \tag{34}$$

where $\widehat{t}^{**}$ solves

$$E_{\widehat{\mu}^*}\left\{\theta(\widehat{\mu}^{**}) - \theta(\widehat{\mu}^*) + \widehat{t}^{**}\right\} = 0 \tag{35}$$

The perturbation $\widetilde{t}$ will improve the approximation and therefore $U(\widehat{\mu}, \widehat{t}^*) = \widehat{t}^* + \widetilde{t}$ is a better approximation for the bias than $\widetilde{t}^*$. Hall's double bootstrap bias corrected estimator is given by,

$$\widetilde{\theta}_{Dbc} = \theta(\widehat{\mu}) + \widehat{t}^* + \widetilde{t} \tag{36}$$

Noting that, $\widehat{t}^{**} = \widehat{H}^{*-1}(0) = \theta(\widehat{\mu})$ (from (35)), $\widehat{t}^* = \widehat{H}^{-1}(0)$, and that $\widetilde{t} = \widehat{H}^{-1}(-\widehat{t}^{**}) = \theta(\widehat{\mu}) - \widehat{G}^{-1}(-\widehat{t}^{**})$, it is easy to find that equation (36) is exactly equal to the expression of double bias corrected estimator in Lemma 2.2.

Basically Hall's method makes the adjustment on $T = \widehat{H}^{-1}(0)$ while Shi's methodology applies the adjustment on $\widehat{H}(\widehat{t}^*)$. Here the two are equal because the function H is linear both through the root function $R_{t(.)}(.,.)$ and through the expectation operator. In the case of confidence limits. Hall's methodology finds the perturbation to the critical value $\widehat{t}$ such that $\widehat{H} = P(\widehat{\theta}^* \leq \widehat{t} + \widetilde{t}) = \alpha$ while Shi's methodology, corrects the probability value that corresponds to $\widehat{t}$ by finding β such that $\widehat{H} = P(\widehat{\theta} \leq \widehat{t}^*) = \alpha$ and $\widehat{H}^* = P(\widehat{\theta}^{**} \leq \widehat{t}^*) = \beta$.