Bootstrap–based criteria for choosing the number of instruments

Okui, R.

Department of Economics, Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong. E-Mail: okui@ust.hk

Keywords: *Instrumental variables estimation; Bootstrap; Instrument selection*

ABSTRACT

Instrumental variables estimation is a widely used technique in many fields in the social sciences. An important problem in applying the technique is the choice of instrumental variables, since empirical researchers often have many (valid) instrumental variables from which to choose. According to conventional asymptotic theory, the asymptotic variance of an estimator decreases with the number of instruments used for estimation. However, instrumental variables estimators behave poorly when there are many moment conditions. In particular, the two-stage least squares (2SLS) estimator, which is the most widely used instrumental variables estimator, has a bias that is proportional to the number of instruments. Appropriate selection of the number of instruments is therefore essential for obtaining good estimates.

The main purpose of this paper is to report the results of Monte Carlo simulations conducted to investigate the ability of bootstrap based criteria for choosing a number of instrumental variables. We consider the 2SLS estimation of a linear simultaneous equations model. The bootstrap is used to estimate the mean square error (MSE) of the estimator, and then the number of instruments is chosen by minimizing the bootstrapped MSE.

The bootstrap is a method for approximating the distribution of a statistic by computing the distribution of that statistic under an estimated distribution of the data (which is often the empirical distribution of the data). In practice, we repeatedly draw samples from an estimated distribution of the data and compute the statistics of interest for each sample. The empirical distributions of the statistics computed in this way are used as estimates of the distributions of the statistics.

This paper considers three different bootstrap procedures. The first is the "naive" bootstrap, in which each bootstrap sample is drawn from the empirical distribution of the data. In the second bootstrap procedure, bootstrap samples are drawn from the empirical distribution, which is modified so that the instruments are orthogonal to the residuals of the equation. The third bootstrap procedure involves an analytical bias correction. When we compute the bootstrapped MSE, we use an analytic approximation of the bias rather than the bootstrap bias estimate.

The results of our experiments show that the "naive" bootstrap does not work for the purpose of choosing instruments. A large number of instruments tend to be chosen by the "naive" bootstrap method. When the sample size is small, the estimator based on the choice of instruments implied by the "naive" bootstrap performs similarly to the estimator based on the use of all available instruments. When the bootstrap sample is drawn so that the instruments are orthogonal to the residuals, the estimator is more precise, but it does not perform well in small samples. The problem is that the bootstrap estimate of the bias is inaccurate. By correcting the bias by using an analytic approximation of the bias rather than the bootstrap bias estimate, the bootstrap–based procedure improves the performance of the estimator. With these modifications, the bootstrap procedure performs equally as well, and often better, than the selection method of Donald and Newey (2001).

1 INTRODUCTION

Empirical researchers can often choose between a large number of instrumental variables. Using many instrumental variables reduces the asymptotic variance of instrumental variables estimators according to conventional asymptotic theory. However, in finite samples, instrumental variables estimators behave poorly when there are many instrumental variables (see, e.g. Kunitomo (1980), Morimune (1983) and Bekker (1994)). It is therefore important to choose the number of instruments appropriately. In this paper, we investigate the effectiveness of bootstrap based criteria for choosing the number of instruments used for estimation.

We consider the following linear simultaneous equations model in which there is a single endogenous regressor but no exogenous regressors:

$$y_i = \delta Y_i + \epsilon_i, \text{ for } i = 1, \dots, N.$$

The variable y_i is a scalar dependent variable and Y_i is a scalar regressor that is correlated with the error term, ϵ_i, so that $E(Y_i \epsilon_i) \neq 0$. We aim to estimate the parameter δ. A set of instruments, Z_i, is available for estimation and we assume that all the instruments are valid: $E(Z_i \epsilon_i) = 0$. Let K denote the number of instruments. For simplicity, we assume conditional homoskedasticity of the errors: $E(\epsilon_i^2) = \sigma^2$ for all i. We also assume that $(y_i, Y_i, Z_i), i = 1, \dots, N$ are i.i.d. Throughout the paper, we use the following notation. For a sequence of vectors, $\{a_i\}_{i=1}^N$, we define a as $a = (a_i, \dots, a_N)'$. For a matrix A, we define P_A as $P_A = A(A'A)^{-1}A'$. For example, the vector Y is defined as $Y = (Y_1, \dots, Y_N)$ and the matrix P_Z is defined as $P_Z = Z(Z'Z)^{-1}Z'$.

Note that extending the model to incorporate multiple endogenous regressors and/or exogenous regressors is straightforward. For ease of exposition, we consider only the model described above.

The most widely used estimator for δ is the 2SLS estimator. The 2SLS estimator that uses Z_i as instruments is defined as

$$\hat{\delta} = (Y'P_Z Y)^{-1} Y' P_Z y.$$

Conventional asymptotic theory implies

$$\sqrt{N}(\hat{\delta} - \delta) \to_d N(0, \sigma^2 H^{-1}), \text{ as } N \to \infty,$$

where $H = E(Y_i Z_i) E(Z_i Z_i')^{-1} E(Z_i Y_i)$ and "$\to_d$" represents "convergence in distribution." It is easy to show that the asymptotic variance decreases if instruments are added. Hence, this theory implies that all available instruments should be used.

However, this recommendation might be inappropriate if there are a large number of instruments. The 2SLS estimator behaves poorly when there are many instruments, primarily because of the associated bias (see, e.g., Kunitomo (1980), Morimune (1983) and Bekker (1994)). The finite sample bias of the 2SLS estimator is approximately

$$E(\hat{\delta} - \delta) \approx \sigma_{\epsilon,u} H^{-1} \frac{K}{N}, \tag{1}$$

where $\sigma_{\epsilon,u} = E(\epsilon_i u_i)$ and u_i is the residual from the regression of the endogenous variable, Y_i, on the sets of instruments, Z_i. See, e.g., Hahn and Hausman (2002a, 2002b) for a derivation of the formula for the bias. The bias of the 2SLS estimator is proportional to the number of instrumental variables. Thus, it is not necessarily desirable to use all available instruments because of the bias–variance trade-off. Hence, a good estimator requires the appropriate selection of instruments.

In this paper, we report the results of Monte Carlo experiments designed to examine the finite sample performance of instrument selection methods based on the bootstrap. The bootstrap is used to estimate the MSE of the 2SLS estimator. We then minimize the estimated MSE to determine the number of instruments used for estimation. The results show that the "naive" bootstrap method is not effective for the purpose of instrument selection. Two modifications are needed. First, the residuals must be made orthogonal to the instruments so that the moment conditions hold under the bootstrap distribution. Second, we use an analytic approximation of the bias rather than the bootstrap bias estimate in computing the MSE. Our simulations reveal that, once these modifications have been made, choosing the number of instruments by using bootstrap–based criteria is effective in improving the precision of the estimator.

Donald and Newey (2001) first proposed choosing the number of instruments by minimizing the estimated MSE. They calculate an analytic approximation of the MSE and use it as a criterion function. Their work has been extended and applied to conditional moment restriction models (Donald, Imbens, and Newey (2002)), time-series models (Kuersteiner (2002)) and dynamic panel data models (Okui (2005)). In our experiments, we compare the performance of bootstrap based procedures with that of the instrument selection method proposed by Donald and Newey (2001). We find that the bootstrap–based method performs as well, and often better, than does their procedure.

Bootstrap–based methods possess several advantages over methods based on analytical approximation of the MSE. For example, the bootstrap enables consideration of criteria other than the MSE, such as the mean absolute deviation. Computing the bootstrapped mean absolute deviation is as easy as computing the bootstrapped MSE. However, it is difficult to approximate the mean absolute deviation of an instrumental variables estimator analytically. A disadvantage of bootstrap–based criteria is their computational cost, although we expect this to diminish over time.

Several studies apply bootstrap methods to address the problem of "many instruments". Inoue (2005) proposes using a bootstrap method to choose the number of instruments. His method minimizes the approximate coverage error of the confidence intervals; the coverage error is estimated by the bootstrap. Flores-Lagunes (2002) proposes a bias-corrected 2SLS estimator in which the bias is estimated by using a bootstrap method, although his main concern is the bias caused by weak instruments. Kleibergen (2004) investigates the applicability of the bootstrap critical values for various test statistics when there are many instruments.

The objective of this paper (and the paper by Donald and Newey (2001)) differs from those on the selection methods proposed by Andrews (1999) and Andrews and Lu (2001). The goal of Andrews (1999) and Andrews and Lu (2001) is to distinguish between correct moment conditions and incorrect ones. In this paper, we assume that all moment conditions are correct, and then select the number of moment conditions to avoid the problem of "many instruments."

The organization of the rest of the paper is as follows. In Section 2, we describe the Monte Carlo design. In Section 3, we explain bootstrap based procedures for choosing the number of instruments. In Section 4, we report the results of the Monte Carlo experiments. Suggestions for future research are offered in Section 5.

2 DESIGN

In this section, we explain the Monte Carlo design, following Donald and Newey (2001). Our data-generating process is the following model:

$$\begin{aligned} y_i &= \delta Y_i + \epsilon_i, \\ Y_i &= \pi' Z_i + u_i, \end{aligned}$$

for $i = 1, \ldots, N$, where Y_i is a scalar, δ is a scalar parameter of interest, $Z_i \sim i.i.d.N(0, I_K)$ and

$$\begin{pmatrix} \epsilon_i \\ u_i \end{pmatrix} \sim i.i.d.N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & c \\ c & 1 \end{pmatrix} \right).$$

The integer K is the total number of instruments. We set the true value of δ as $\delta = 0.1$, and then examine how well each estimator estimates δ.

In this framework, each experiment is indexed by the vector of specifications, (N, K, c, π').We use $N = 100$ and $N = 500$, and set $K = 20$ if $N = 100$ and set $K = 25$ if $N = 500$. The degree of endogeneity is summarized by c, which is set to $c = 0.1, 0.5$ and 0.9.

We consider the specification of the vector π used in Donald and Newey (2001). Let R_f^2 denote the theoretical R^2 of the first-stage regression. The k–th element of π is

$$\pi_k = c(K) \left(1 - \frac{k}{K+1} \right)^4,$$

where $c(K)$ is chosen to satisfy $\pi'\pi = R_f^2/(1 - R_f^2)$. Note that $R_f^2 = \pi'\pi/(\pi'\pi + 1)$ (Hahn and Hausman (2002a)). We use $R_f^2 = 0.1$ and 0.2. The strength of the instruments decreases moderately in k. We perform an experiment by assuming that the rank ordering of the instruments is known. That is, the instruments are ordered according to their strengths, and we choose only the number of instruments used for estimation.

We also tried other specifications of π. The results of the experiments based on other specifications are available from the author on request. The relative performance of each estimator is robust to the specification of π.

3 PROCEDURES

In this section, we describe the procedures in relation to which estimation performance is examined in the Monte Carlo simulations. As benchmarks, the 2SLS estimator, which uses all the available instruments ("2SLS" in the tables), and the 2SLS estimator, which uses the number of instruments chosen on the basis of the procedure of Donald and Newey (2001) ("DN" in the tables), are examined in the experiments. Three bootstrap–based procedures are explained in detail.

The bootstrap is a way of approximating the distribution of a statistic by computing the distribution of the statistic based on an estimated distribution of the data. We use bootstrap methods for approximating the MSE of the 2SLS estimator, and then use the bootstrapped MSE as a criterion function for choosing the number of instruments.

First, we consider the "naive" bootstrap, in which each bootstrap sample is drawn from the empirical distribution of the data.

Algorithm 1 (B1). *1. Draw the bootstrap sample, $\{y_i^*, Y_i^*, Z_i^*\}_{i=1}^N$, from the empirical distribution of the data, $\{y_i, Y_i, Z_i\}_{i=1}^N$.*

2. Compute the 2SLS estimate for the bootstrap sample: $\hat{\delta}^*(k) = (Y^{*\prime} P_{z(k)^*} Y^*)^{-1} Y^{*\prime} P_{z(k)^*} y^*$, for each k.

The vector $z_i(k)$ is the vector containing the first k elements of the vector Z_i^*.

3. *Repeat Steps 1 and 2 B times.*
4. *Compute the following bootstrapped MSE:*

$$\text{BMSE}(k) = \frac{1}{B}\sum_{b=1}^{B}(\hat{\delta}_b^*(k) - \hat{\delta}(k))^2,$$

where $\hat{\delta}_b^*(k)$ is the 2SLS estimator that uses $kinstruments$for the b–th bootstrap sample and $\hat{\delta}(k)$ is the 2SLS estimator that uses k instruments.

5. *Choose the number of instruments that minimizes BMSE(k).*

Hahn (1996) shows that the asymptotic distribution of $\hat{\delta}_b^*(k)$ is the same as that of $\hat{\delta}$: $\sqrt{N}(\hat{\delta}_b^*(k) - \hat{\delta}(k)) \to_d N(0, \sigma^2 H^{-1})$. However, as shown by the simulations, this bootstrap procedure is not effective for instrument selection. The "naive" bootstrap needs to be modified in two ways. First, the moment conditions are re-centered so that they hold under the bootstrap distribution. Second, we replace the bootstrap estimate of the bias by a bias estimate based on an analytic approximation of the bias.

Consider the first modification. The moment conditions do not hold under the empirical distribution in general. This is because, in general, there is no value of δ that satisfies $\sum_{i=1}^{N} z_i(k)(y_i - Y_i\delta)/N = 0$, when the number of instruments is larger than the number of parameters. The empirical distribution therefore fails to provide a good approximation of the current model. To solve this problem, we consider the following bootstrap procedure, in which each bootstrap sample is drawn from a distribution that satisfies the moment conditions.

Algorithm 2 (B2). *1. Let $\hat{\delta}$ be a consistent estimate of δ.*

2. *Let $\tilde{\epsilon} = (I - P_Z)(y - \hat{\delta}Y)$. Draw the bootstrap sample, $\{Y_i^*, Z_i^*, \epsilon_i^*\}_{i=1}^N$ from the empirical distribution of $\{Y_i, Z_i, \tilde{\epsilon}_i\}_{i=1}^N$. Let $y_i^* = \hat{\delta}Y_i^* + \epsilon_i^*$.*
3. *Compute the 2SLS estimates for the bootstrap sample: $\hat{\delta}^*(k) = (Y^{*\prime}P_{z(k)^*}Y^*)^{-1}Y^{*\prime}P_{z(k)^*}y^*$ for $k = 1, \ldots, K$. The vector $z_i(k)$ is the vector containing the first k elements of the vector, Z_i^*.*
4. *Repeat Steps 2 and 3 B times.*
5. *Compute the following bootstrapped MSE:*

$$\text{BMSE}(k) = \frac{1}{B}\sum_{b=1}^{B}(\hat{\delta}_b^*(k) - \hat{\delta})^2,$$

where $\hat{\delta}_b^*(k)$ is the 2SLS estimator that uses k instruments for the b–th bootstrap sample.

6. *Choose the number of instruments that minimizes BMSE(k).*

In the experiments, the preliminary estimate, $\hat{\delta}$, is obtained by using the number of instruments that minimizes the first-stage cross-validation criteria, as in Donald and Newey (2001).

In Step 2, we multiply the matrix $(I - P_Z)$ by the vector of residuals, $(y - \hat{\delta}Y)$. The transformed residuals are orthogonal to the instruments: $\sum_{i=1}^{N} Z_i\tilde{\epsilon}_i/N = 0$. This process therefore re-centers the moment conditions, so that the moment conditions hold under the bootstrap distribution. Several studies demonstrate the importance of re-centering of the moment conditions (see, e.g., Horowitz (2002, Section 3.7) and Hall and Horowitz (1996)). For example, re-centering is needed for asymptotic refinements to the size of the t test and for the coverage probability of a confidence interval. Moreover, it is important to re-center the moment conditions to estimate the distribution of the test statistic for over-identifying restrictions by the bootstrap.

It is also important to use the same bootstrap sample to compute the MSEs of 2SLS estimators with different sets of instruments. When we compute the bootstrapped MSE of the 2SLS estimator that uses k instruments, we could consider using the residuals that are orthogonal to only k instruments. This process, however, causes a problem since different data sets are used to compute the MSEs of different estimators. This makes the comparison of estimators difficult.

The estimate for the bootstrap sample, $\hat{\delta}_b^*$, is centered at $\hat{\delta}$ here. Note that $\hat{\delta}$ is the true value of δ under the bootstrap distribution, since $\hat{\delta}$ satisfies the moment conditions under the bootstrap distribution.

Theorem 3.1 of Freedman (1984) provides a theoretical justification for the bootstrap scheme, "B2." It shows that the distance between the distribution of $\sqrt{N}(\hat{\delta}^*(k) - \hat{\delta})$ and the distribution of $\sqrt{N}(\hat{\delta}(k) - \delta)$ approaches zero asymptotically. This property is known as the bootstrap principle.

Lastly, we consider the bootstrap procedure in which the bias of the estimator is corrected "manually." It has been observed that the bootstrap is not always effective in measuring the bias of an estimator. As shown in the next section, this problem arises in the context of our experiments. To mitigate this problem, we replace the bootstrap estimate of the bias by a bias estimate based on an analytic approximation of the bias. This strategy is also used by Härdle and Bowman (1988) and Nishiyama and Robinson (2005) in different contexts. The bootstrap procedure based on this modification is summarized as Algorithm 3.

Algorithm 3 (B3). *1. Let $\hat{\delta}$ be a consistent estimate of δ.*

2. *Let $\tilde{\epsilon} = (I - P_Z)(y - \hat{\delta}Y)$. Draw the bootstrap sample, $\{Y_i^*, Z_i^*, \epsilon_i^*\}_{i=1}^N$ from the empirical distribution of $\{Y_i, Z_i, \tilde{\epsilon}_i\}_{i=1}^N$. Let $y_i^* = \hat{\delta}Y_i^* + \epsilon_i^*$.*

3. *Compute the following 2SLS estimates for the bootstrap sample: $\hat{\delta}^*(k) = (Y^{*\prime}P_{z(k)^*}Y^*)^{-1}Y^{*\prime}P_{z(k)^*}y^*$ for $k = 1, \ldots, K$*

4. *Repeat Steps 2 and 3 B times.*

5. *Compute the following bootstrapped MSE:*

$$BMSE(k) = \frac{1}{B}\sum_{b=1}^{B}(\hat{\delta}_b^*(k) - \bar{\delta}^*(k) + \hat{C}(k))^2,$$

where $\hat{\delta}_b^*(k)$ is the 2SLS estimator that uses k instruments for the b–th bootstrap sample, $\bar{\delta}^*(k) = \sum_{b=1}^{B}\hat{\delta}_b^*(k)/B$, $\hat{C}(k) = (Y'P_ZY)^{-1}K\hat{\sigma}_{\epsilon,u}$, and $\hat{\sigma}_{\epsilon,u} = \sum_{i=1}^{N}(y_i - \hat{\delta}Y_i)(Y_i - z_i(Z'Z)^{-1}Z'Y)/N$.

6. *Choose the number of instruments that minimizes BMSE(k).*

The only difference between B2 and B3 arises in Step 5. Note that

$$\begin{aligned} & \frac{1}{B}\sum_{b=1}^{B}(\hat{\delta}_b^*(k) - \hat{\delta})^2 \\ = \quad & \frac{1}{B}\sum_{b=1}^{B}(\hat{\delta}_b^*(k) - \bar{\delta}^*(k))^2 + (\bar{\delta}^*(k) - \hat{\delta})^2, \end{aligned}$$

where the first term on the right-hand side is the bootstrap estimate of the variance and the second term corresponds to the square of the bootstrap bias estimate. We replace the bootstrap bias estimate, $\bar{\delta}^*(k) - \hat{\delta}$, by an estimate of the bias based on the analytic approximation, $\hat{C}(k)$ (c.f. formula (1) in Section 2).

It is not difficult to show that bootstrap procedure B3 also satisfies the bootstrap principle. A proof similar to Theorem 3.1 in Freedman (1984) could be used.

One advantage of the bootstrap is that it enables the use of many different criteria. For example, the algorithm for the bootstrapped mean absolute deviation criterion is similar to Algorithm 3 except that we replace "BMSE(k)" in Step 5 by

$$\text{BA}(k) = \frac{1}{B}\sum_{b=1}^{B}\left|\hat{\delta}_b(k) - \bar{\delta}^*(k) + \hat{C}(k)\right|.$$

Table 1. Monte Carlo results: $R^2 = 0.1$

	2SLS	DN	B1	B2	B3
			$c = 0.1, N = 100$		
BIAS	0.0586	0.031	0.0655	0.0575	0.0398
MAD	0.134	0.201	0.126	0.14	0.177
CR	0.95	0.972	0.962	0.946	0.964
			$c = 0.1, N = 500$		
BIAS	0.0306	0.0151	0.0329	0.0302	0.0246
MAD	0.076	0.0881	0.757	0.0779	0.0844
CR	0.961	0.954	0.967	0.956	0.949
			$c = 0.5, N = 100$		
BIAS	0.307	0.153	0.317	0.283	0.154
MAD	0.308	0.245	0.318	0.291	0.237
CR	0.51	0.833	0.5	0.586	0.805
			$c = 0.5, N = 500$		
BIAS	0.15	0.0663	0.138	0.0929	0.0716
MAD	0.151	0.1	0.146	0.112	0.104
CR	0.676	0.895	0.711	0.817	0.887
			$c = 0.9, N = 100$		
BIAS	0.566	0.235	0.565	0.378	0.26
MAD	0.566	0.324	0.565	0.378	0.295
CR	0.017	0.749	0.047	0.41	0.675
			$c = 0.9, N = 500$		
BIAS	0.273	0.0975	0.15	0.114	0.101
MAD	0.273	0.126	0.153	0.128	0.116
CR	0.168	0.844	0.657	0.784	0.829

On the other hand, it is difficult to calculate an analytic approximation of some quantities, such as the mean absolute deviation.

4 RESULTS

In this section, we report the results of the Monte Carlo experiments. The experiments are performed by using Ox 3.40 (Doornik (2002)) for Linux. The number of Monte Carlo replications is set to 1000 and the number of bootstrap replications is set to 500 in all experiments.

Tables 1 and 2 report the results of the experiments. For each estimator, we compute the median bias (BIAS), the median absolute deviation (MAD) and the coverage rate (CR) for a 95% (Wald-type) confidence interval. We use these "robust" measures because of concerns about the existence of moments of estimators. The measure "MAD" would best represent the performance of each estimator. The magnitude of "BIAS" indicates the number of instruments chosen by each procedure. Roughly, a large "BIAS" implies a large number of instruments.

B1 suggests a large number of instruments. It produces a good estimate when $c = 0.1$, in which case, many instruments should be used, but yields a poor estimate when $c = 0.5$ or 0.9, in which case, the optimal number of instruments is small. For cases

Table 2. Monte Carlo results: $R^2 = 0.2$

	2SLS	DN	B1	B2	B3
			$c = 0.1$, $N = 100$		
BIAS	0.0422	0.0166	0.0458	0.0412	0.0241
MAD	0.111	0.13	0.104	0.113	0.122
CR	0.949	0.957	0.962	0.941	0.949
			$c = 0.1$, $N = 500$		
BIAS	0.0161	0.0094	0.0176	0.0151	0.0134
MAD	0.0545	0.0587	0.0547	0.055	0.057
CR	0.959	0.949	0.959	0.959	0.952
			$c = 0.5$, $N = 100$		
BIAS	0.209	0.0992	0.21	0.169	0.0986
MAD	0.209	0.157	0.212	0.186	0.157
CR	0.641	0.874	0.642	0.729	0.845
			$c = 0.5$, $N = 500$		
BIAS	0.0806	0.0342	0.0524	0.0445	0.0382
MAD	0.0861	0.064	0.0695	0.0671	0.0654
CR	0.81	0.928	0.877	0.905	0.922
			$c = 0.9$, $N = 100$		
BIAS	0.384	0.134	0.342	0.191	0.144
MAD	0.384	0.186	0.342	0.202	0.173
CR	0.104	0.824	0.29	0.679	0.796
			$c = 0.9$, $N = 500$		
BIAS	0.145	0.0525	0.0695	0.0595	0.0528
MAD	0.145	0.0737	0.0796	0.0732	0.0702
CR	0.475	0.889	0.841	0.875	0.886

with $N = 100$ and $R^2 = 0.1$, the performance of B1 is similar to that of 2SLS. For cases with $N = 500$ and $R^2 = 0.2$, B1 yields a lower median absolute deviation of the estimator. The use of procedure B1 improves the performance of the 2SLS estimator only if there is a large sample and/or strong instruments. Since the problem of "many instruments" is serious in small samples, a procedure that needs a large sample to solve the problem is not particularly useful. Moreover, even when the use of B1 improves the performance of the estimator, the improvement is small relative to that achieved by other instruments selection procedures.

Procedure B2 provides more precise estimates than does 2SLS (at least in terms of the median absolute deviation). In particular, B2 performs well when $N = 500$. DN and B2 perform similarly when $N = 500$ and $c = 0.5$ or 0.9, and B2 outperforms DN when $N = 500$ and $c = 0.1$. When the sample is small ($N = 100$) and the degree of endogeneity is moderate or high ($c = 0.5$ or 0.9), the number of instruments chosen on the basis of B2 is above the optimal number and B2 does not perform well when compared with DN and B3. When $c = 0.1$, B2 outperforms DN, but this seems to be because B2 tends to imply that a relatively large number of instruments is chosen.

B2 performs poorly because the bootstrap estimate of the bias is inaccurate. Table 3 compares the 2SLS bias with its bootstrap estimate when $R^2 = 0.1$. In

Table 3. The bias of the 2SLS estimator and the bootstrap estimate of the bias: $R^2 = 0.1$

c	N	2SLS	B2
0.1	100	0.0568	-0.01
	500	0.028	0.02
0.5	100	0.308	0.126
	500	0.148	0.0979
0.9	100	0.567	0.242
	500	0.269	0.175

the columns "2SLS" and "B2" are the actual bias of the 2SLS estimator, which uses all the available instruments, and the mean of the bootstrap estimate of the bias, respectively. Table 3 shows that the bootstrap underestimates the bias that is due to the inclusion of many instruments. This property of the bootstrap explains why B2 implies that a relatively large number of instruments is chosen.

Using procedure B3 improves the precision of the estimator substantially when $c = 0.5$ and 0.9. When $c = 0.1$, B2 outperforms B3, but B3 outperforms DN. Note that instrument selection is important when there is a high degree of endogeneity (see formula (1) in Section 2). It is worth noting that B3 produces a smaller median absolute deviation than does DN in most cases.

To summarize, we find that the "naive" bootstrap is ineffective in choosing the number of instruments. Use of the "naive" bootstrap tends to lead to a large number of instruments being chosen. When the sample is small, the performance of the estimator based on the choice of instruments implied by the "naive" bootstrap is similar to that of the estimator based on the use of all available instruments. If the bootstrap sample is drawn so that the moment conditions hold under the bootstrap distribution, the estimator is more precise, but it does not perform well in small samples. The problem is that the bootstrap underestimates the bias caused by the use of many instruments. When the bias is corrected by using an analytic approximation of the bias rather than the bootstrap bias estimate, the procedure's performance is satisfactory. With these modifications, the bootstrap procedure performs as well, and often better, than does the selection method of Donald and Newey (2001).

5 DISCUSSION

In this paper, we investigate the effectiveness of bootstrap based criteria for choosing the number of instruments used for estimation by Monte Carlo simulations. The results are encouraging and should stimulate further research on bootstrap based procedures. In particular, two problems seem important.

First, a theoretical justification for using the bootstrap procedure examined in this paper is provided by an argument based on the theory of Freedman (1984), in which the number of instruments is assumed to be fixed. However, the choice of instruments is important when there are many instruments. In this case, double asymptotics, under which both the number of instruments and the sample size approach infinity, would provide a better approximation. Theoretical research on the properties of the bootstrap under double asymptotics is therefore desirable.

Second, there are many other situations in which the problem of many moment conditions is important. For example, dynamic panel data models typically yield a large number of moment conditions. Hence, it would be interesting to investigate the effectiveness of bootstrap based criteria in other situations.

6 ACKNOWLEDGEMENT

The author thanks Atsushi Inoue for helpful discussions. The author is responsible for all errors.

7 REFERENCES

Andrews, D. W. (1999), Consistent moment selection procedures for generalized method of moments estimation, *Econometrica*, **67(3)**, 543–564.

Andrews, D. W. and Lu, B. (2001), Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models, *Journal of Econometrics*, **101**, 123–164.

Bekker, P. A. (1994), Alternative approximations to the distributions of instrumental variable estimators, *Econometrica*, **62(3)**, 657–681.

Donald, S. G., Imbens, G. W., and Newey, W. K. (2002), Choosing the number of moments in conditional moment restriction models, mimeo.

Donald, S. G. and Newey, W. K. (2001), Choosing the number of instruments, *Econometrica*, **67(5)**, 1161–1191.

Doornik, J. A. (2002) *Object-Oriented Matrix Programming Using Ox*, 3rd ed. London: Timberlake Consultants Press and Oxford: www.nuff.ox.ac.uk/Users/Doornik.

Flores-Lagunes, A. (2002), Estimation methods robust to weak instruments, mimeo.

Freedman, D. (1984), On bootstrapping two-stage least-squares estimates in stationary linear models, *Annals of Statistics*, **12(3)**, 827–842.

Hahn, J. (1996), A note on bootstrapping generalized method of moments estimators, *Econometric Theory*, **12**, 187–197.

Hahn, J. and Hausman, J. (2002a), A new specification test for the validity of instrumental variables, *Econometrica*, **70(1)**, 163–189.

Hahn, J. and Hausman, J. (2002b), Notes on bias in estimators for simultaneous equation models, *Econometric Letters*, **75)**, 237–241.

Hall, P. and Horowitz J. L. (1996), Bootstrap critical values for tests based on generalized–method–of–moments estimators, *Econometrica*, **64(4)**, 891–916.

Härdle, W. and Bowman, A. W. (1988), Bootstrapping in nonparametric regression: local adaptive smoothing and confidence bands, *Journal of American Statistical Association*, **83(401)**, 102–110.

Horowitz, J. L. (2002) The bootstrap, in J. J. Heckman and E. Leamer, eds., *Handbook of Econometrics*, vol. 5, chap. 52, 3159–3228, Elsevier Science B. V.

Inoue, A. (2005), A bootstrap approach to moment selection, mimeo.

Kleibergen, F. (2004), Expansions of GMM statistics that indicate their properties under weak and/or many instruments and the bootstrap, mimeo.

Kuersteiner, G. (2002), Rate-adaptive GMM estimators for linear time-series models, mimeo.

Kunitomo, N. (1980), Asymptotic expansions of the distributions of estimators in a linear functional relationship and simultaneous equations, *Journal of the American Statistical Association*, **75**, 693–700.

Morimune, K. (1983), Approximate distributions of k-class estimators when the degree of overidentification is large compared with the sample size, *Econometrica*, **51(3)**, 821–841.

Nishiyama, Y. and Robinson, P. M. (2005), The bootstrap and the Edgeworth correction for semiparametric averaged derivatives, *Econometrica*, **73(3)**, 903–948.

Okui, R. (2005), The optimal choice of moments in dynamic panel data models, mimeo.