ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Davidson, Russell; MacKinnon, James G.

Working Paper

Bootstrap confidence sets with weak instruments

Queen's Economics Department Working Paper, No. 1278

Provided in Cooperation with:
Queen's University, Department of Economics (QED)

Suggested Citation: Davidson, Russell; MacKinnon, James G. (2011) : Bootstrap confidence sets with weak instruments, Queen's Economics Department Working Paper, No. 1278, Queen's University, Department of Economics, Kingston (Ontario)

This Version is available at:
https://hdl.handle.net/10419/67786

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Queen's Economics Department Working Paper No. 1278

Bootstrap Confidence Sets with Weak Instruments

Russell Davidson
McGill University

James G. MacKinnon
Queen's University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

10-2011

Bootstrap Confidence Sets with Weak Instruments

by

Russell Davidson

Department of Economics and CIREQ
McGill University
Montréal, Québec, Canada
H3A 2T7

GREQAM
Centre de la Vieille Charité
2 Rue de la Charité
13236 Marseille cedex 02, France

russell.davidson@mcgill.ca

and

James G. MacKinnon

Department of Economics
Queen's University
Kingston, Ontario, Canada
K7L 3N6

jgm@econ.queensu.ca

Abstract

We study several methods of constructing confidence sets for the coefficient of the single right-hand-side endogenous variable in a linear equation with weak instruments. Two of these are based on conditional likelihood ratio (CLR) tests, and the others are based on inverting t statistics or the bootstrap P values associated with them. We propose a new method for constructing bootstrap confidence sets based on t statistics. In large samples, the procedures that generally work best are CLR confidence sets using asymptotic critical values and bootstrap confidence sets based on LIML estimates.

Key words: Weak instruments, bootstrap, confidence sets

JEL codes: C10, C15

This research was supported, in part, by grants from the Social Sciences and Humanities Research Council of Canada, the Canada Research Chairs program (Chair in Economics, McGill University), and the Fonds Québécois de Recherche sur la Société et la Culture.

October 2, 2011.

1. Introduction

Inference in the linear simultaneous equations model is notoriously difficult when the instruments are weak. Although there has been an enormous amount of work on this topic since the seminal paper of Staiger and Stock (1997), much of it has focused on the properties of estimators (especially their bias) and on the properties of test statistics. Despite important work by Zivot, Startz, and Nelson (1998), Mikusheva (2010),[1] and many others, there does not yet appear to be a consensus on the best way to construct confidence sets when instruments are weak. This paper examines several procedures that are either easy to use and popular or may be expected to perform well. We obtain a number of striking results.

In principle, one can construct a confidence set by inverting any suitable test statistic, possibly after it has been bootstrapped in some way. For the linear simultaneous equations model, the natural candidates are Wald (that is, t) tests, likelihood ratio (LR) tests, Lagrange multiplier (LM) tests, and the Anderson-Rubin (AR) test.

Partly for reasons of space and readability, we restrict attention to confidence sets that are based on Wald tests or on the conditional LR (CLR) test of Moreira (2003). We consider Wald-based confidence sets because they are the most commonly used in practice and because, contrary to what is widely believed, it is possible to make them perform well when the instruments are weak by using certain bootstrap methods. We consider CLR confidence sets because the CLR test often seems to perform very well and because the results of Mikusheva (2010) suggest that CLR confidence sets also perform well.

We do not consider confidence sets based on the LM test or the closely related test of Kleibergen (2002) because the results of Mikusheva (2010) are not at all encouraging. It is partly for the same reason that we do not consider confidence sets based on the AR test of Anderson and Rubin (1949). More importantly, as was shown in Davidson and MacKinnon (2011), AR confidence sets have many undesirable properties. Although their unconditional coverage is, under classical assumptions, always correct, their coverage conditional on being bounded intervals can be far from correct. Moreover, the lengths of AR intervals, when they exist, provide grossly unreliable information about the precision with which the parameter of interest has been estimated.

In the next section, we discuss the basic model and some conventional procedures, both asymptotic and bootstrap, for constructing Wald-based confidence intervals. In Section 3, we discuss confidence sets based on the CLR test. In Section 4, we discuss a new procedure for constructing Wald-based bootstrap confidence intervals. In Section 5, we present a number of simulation results, some of which may be quite surprising. In Section 6, we summarize our conclusions.

---

[1] We are grateful to Lynda Khalaf for drawing our attention to this paper.

2. Wald-Based Confidence Intervals

We restrict attention to the two-equation linear model

$$\boldsymbol{y}_1 = \beta \boldsymbol{y}_2 + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{u}_1 \tag{1}$$
$$\boldsymbol{y}_2 = \boldsymbol{W}\boldsymbol{\pi} + \boldsymbol{u}_2 = \boldsymbol{Z}\boldsymbol{\pi}_1 + \boldsymbol{W}_2\boldsymbol{\pi}_2 + \boldsymbol{u}_2. \tag{2}$$

Here $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are n–vectors of observations on endogenous variables, $\boldsymbol{Z}$ is an $n \times k$ matrix of observations on exogenous variables, and $\boldsymbol{W}$ is an $n \times l$ matrix of exogenous instruments with the property that $\mathcal{S}(\boldsymbol{Z})$, the subspace spanned by the columns of $\boldsymbol{Z}$, lies in $\mathcal{S}(\boldsymbol{W})$, the subspace spanned by the columns of $\boldsymbol{W}$. The $n \times (l-k)$ matrix $\boldsymbol{W}_2$ is constructed in such a way that $\mathcal{S}(\boldsymbol{Z}, \boldsymbol{W}_2) = \mathcal{S}(\boldsymbol{W})$. Equation (1) is a structural equation, and equation (2) is a reduced-form equation. The parameter of interest is β, the coefficient on $\boldsymbol{y}_2$ in equation (1).

The disturbance vectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ are assumed to be serially uncorrelated and homoskedastic, with mean zero and contemporaneous covariance matrix

$$\boldsymbol{\Sigma} \equiv \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

We assume that the model is either exactly identified or overidentified, which implies that $l \geq k+1$. The number of overidentifying restrictions is $l-k-1$.

Equations (1) and (2) can be estimated in many ways. We restrict attention to the two most common single-equation methods, namely, generalized instrumental variables (IV), which is numerically identical to two-stage least squares, and limited-information maximum likelihood (LIML). The two estimators of β are, in self-evident notation, $\hat{\beta}_{\text{IV}}$ and $\hat{\beta}_{\text{LIML}}$, and their standard errors are $\hat{s}_{\text{IV}}$ and $\hat{s}_{\text{LIML}}$.

The simplest and most natural way to form a confidence interval for β in (1) is to invert the t statistic for $\beta = \beta_0$, which is the signed square root of the Wald statistic. This yields the asymptotic Wald intervals

$$[\hat{\beta}_{\text{IV}} - \Phi_{1-\alpha/2}\hat{s}_{\text{IV}},\ \ \hat{\beta}_{\text{IV}} + \Phi_{1-\alpha/2}\hat{s}_{\text{IV}}] \tag{3}$$

and

$$[\hat{\beta}_{\text{LIML}} - \Phi_{1-\alpha/2}\hat{s}_{\text{LIML}},\ \ \hat{\beta}_{\text{LIML}} + \Phi_{1-\alpha/2}\hat{s}_{\text{LIML}}], \tag{4}$$

where $\Phi_{1-\alpha/2}$ denotes the $1-\alpha/2$ quantile of the standard normal distribution. However, as is well-known and will be seen again in Section 5, these intervals often have poor finite-sample properties when the instruments are weak. This is particularly true for (3), in part because $\hat{\beta}_{\text{IV}}$ can be severely biased in that case.

A natural way in which to attempt to obtain more reliable Wald intervals is to use the bootstrap. The oldest, and conceptually the simplest, bootstrap method for the linear simultaneous equations model is the pairs bootstrap, which was proposed by Freedman (1984). The idea is simply to resample the rows of the matrix $[\boldsymbol{y}_1\ \ \boldsymbol{y}_2\ \ \boldsymbol{Z}\ \ \boldsymbol{W}_2]$.

Each such bootstrap sample, indexed by $j = 1, \ldots, B$, is then used to compute a bootstrap t statistic

$$t_j^* = \frac{\hat{\beta}_j^* - \hat{\beta}}{s(\hat{\beta}_j^*)},$$

where $\hat{\beta}$ could be either $\hat{\beta}_{\text{IV}}$ or $\hat{\beta}_{\text{LIML}}$, $\hat{\beta}_j^*$ is the corresponding estimate from the j^{th} bootstrap sample, and $s(\hat{\beta}_j^*)$ is the standard error of $\hat{\beta}_j^*$. Using either $\hat{\beta}_{\text{IV}}$ and $\hat{s}_{\text{IV}}$ or $\hat{\beta}_{\text{LIML}}$ and $\hat{s}_{\text{LIML}}$, together with the B values of t_j^*, one then constructs an equal-tail percentile t confidence interval (also called a studentized bootstrap confidence interval) in the usual way; see, among many others, Davison and Hinkley (1997) or Davidson and MacKinnon (2004, Chapter 5). In the IV case, the interval is

$$[\hat{\beta}_{\text{IV}} - c_{1-\alpha/2}^* \hat{s}_{\text{IV}}, \;\; \hat{\beta}_{\text{IV}} - c_{\alpha/2}^* \hat{s}_{\text{IV}}], \tag{5}$$

where $c_{\alpha/2}^*$ and $c_{1-\alpha/2}^*$ denote the estimated $\alpha/2$ and $1-\alpha/2$ quantiles of the t_j^*. When $B = 999$ and $\alpha = 0.05$, for example, these are just numbers 25 and 975 in the list of the t_j^* sorted from smallest to largest.

The Wald-based intervals (3), (4), and (5) are easy to construct and commonly used, but they cannot possibly have correct coverage when the instruments are weak, because they cannot be unbounded. When the instruments in a linear simultaneous-equations model are sufficiently weak, a confidence set with correct coverage must be unbounded with positive probability; see Gleser and Hwang (1987) and Dufour (1997). Unlike these Wald-based intervals, the confidence sets discussed in the next two sections can be unbounded with positive probability.

3. CLR Confidence Sets

Because the CLR test of Moreira (2003) seems to work better than other asymptotic tests for the value of β, it is natural to consider confidence sets obtained by inverting CLR tests. Mikusheva (2010) discusses confidence sets of this type. In this section, we present a different derivation which emphasizes computational issues.

The CLR test statistic and all associated quantities, including $\hat{\beta}_{\text{IV}}$, $\hat{\beta}_{\text{LIML}}$, and their standard errors, depend on the data only through the six quantities

$$\begin{aligned} &P_{11} \equiv \boldsymbol{y}_1^\top \boldsymbol{P}_1 \boldsymbol{y}_1, \quad P_{12} \equiv \boldsymbol{y}_1^\top \boldsymbol{P}_1 \boldsymbol{y}_2, \quad P_{22} \equiv \boldsymbol{y}_2^\top \boldsymbol{P}_1 \boldsymbol{y}_2, \\ &M_{11} \equiv \boldsymbol{y}_1^\top \boldsymbol{M}_{\boldsymbol{W}} \boldsymbol{y}_1, \quad M_{12} \equiv \boldsymbol{y}_1^\top \boldsymbol{M}_{\boldsymbol{W}} \boldsymbol{y}_2, \text{ and } M_{22} \equiv \boldsymbol{y}_2^\top \boldsymbol{M}_{\boldsymbol{W}} \boldsymbol{y}_2, \end{aligned} \tag{6}$$

where $\boldsymbol{M}_{\boldsymbol{W}} \equiv \mathbf{I} - \boldsymbol{W}(\boldsymbol{W}^\top \boldsymbol{W})^{-1} \boldsymbol{W}^\top$, $\boldsymbol{P}_1 \equiv \boldsymbol{M}_{\boldsymbol{Z}} - \boldsymbol{M}_{\boldsymbol{W}}$, and $\boldsymbol{M}_{\boldsymbol{Z}} \equiv \mathbf{I} - \boldsymbol{Z}(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \boldsymbol{Z}^\top$. These six quantities just depend on sums of squared residuals and/or sums of cross-products of residuals from the regressions of $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ on $\boldsymbol{Z}$ and $\boldsymbol{W}$.

In order to compute the CLR test statistic for the hypothesis that $\beta = \beta_0$, we also need the quantities

$$\begin{aligned} &Q_{11} \equiv P_{11} - 2\beta_0 P_{12} + \beta_0^2 P_{22}, \quad Q_{12} \equiv P_{12} - \beta_0 P_{22}, \quad Q_{22} \equiv P_{22}, \\ &N_{11} \equiv M_{11} - 2\beta_0 M_{12} + \beta_0^2 M_{22}, \quad N_{12} \equiv M_{12} - \beta_0 M_{22}, \text{ and } N_{22} \equiv M_{22}. \end{aligned} \tag{7}$$

From these, we calculate

$$SS(\beta_0) \equiv nQ_{11}/N_{11}, \tag{8}$$

$$ST(\beta_0) \equiv \frac{n}{\Delta^{1/2}}\left(Q_{12} - \frac{Q_{11}N_{12}}{N_{11}}\right), \text{ and} \tag{9}$$

$$TT(\beta_0) \equiv \frac{n}{\Delta}\left(Q_{22}N_{11} - 2Q_{12}N_{12} + \frac{Q_{11}N_{12}^2}{N_{11}}\right), \tag{10}$$

where

$$\Delta \equiv N_{11}N_{22} - N_{12}^2 = M_{11}M_{22} - M_{12}^2. \tag{11}$$

It is easy to verify that Δ does not depend on β_0. In Mikusheva (2010), it is shown that the eigenvalues of the 2×2 matrix

$$\begin{bmatrix} SS(\beta_0) & ST(\beta_0) \\ ST(\beta_0) & TT(\beta_0) \end{bmatrix}$$

also do not depend on β_0. These eigenvalues are

$$\tfrac{1}{2}\left(SS(\beta_0) + TT(\beta_0) \pm \sqrt{\big(SS(\beta_0) - TT(\beta_0)\big)^2 + 4ST^2(\beta_0)}\right).$$

It follows that $I_1 \equiv SS(\beta_0) + TT(\beta_0)$ and $I_2 \equiv \big(SS(\beta_0) - TT(\beta_0)\big)^2 + 4ST^2(\beta_0)$ are also independent of β_0.

The LR statistic for testing the hypothesis that $\beta = \beta_0$ takes the form

$$\mathrm{LR}(\beta_0) = n\,\log\big(1 + SS(\beta_0)/n\big) - n\,\log\big(1 + (I_1 - I_2)/2n\big); \tag{12}$$

see, among others, Davidson and MacKinnon (2008). The LR statistic depends on β_0 only through $SS(\beta_0)$. The concentrated loglikelihood function for model (3) is a deterministic, decreasing, function of $SS(\beta_0)$. It is therefore maximized by minimizing $SS(\beta_0)$, for which the minimizer is $\hat{\beta}_{\mathrm{LIML}}$. It follows that the LR statistic is also minimized at $\beta_0 = \hat{\beta}_{\mathrm{LIML}}$ and that $\mathrm{LR}(\hat{\beta}_{\mathrm{LIML}}) = 0$.

Following Moreira (2003) and Mikusheva (2010), the LR statistic (12) can be simplified by Taylor expanding the logarithms and discarding terms that tend to zero as $n \to \infty$. This yields

$$\mathrm{LR}_0(\beta_0) = \tfrac{1}{2}\big(SS(\beta_0) - TT(\beta_0) + I_2\big) = M - TT(\beta_0), \tag{13}$$

where $M \equiv \frac{1}{2}(I_1+I_2)$. The rightmost expression in (13) tells us that $TT(\hat{\beta}_{\mathrm{LIML}}) = M$ and that $\mathrm{LR}_0(\hat{\beta}_{\mathrm{LIML}}) = 0$. This implies that $\hat{\beta}_{\mathrm{LIML}}$ belongs to any confidence set found by inverting the test based on $\mathrm{LR}_0(\beta_0)$.

The idea behind the CLR test is that, even though $\mathrm{LR}_0(\beta_0)$ is not pivotal with weak instruments, its distribution conditional on $TT(\beta_0)$ is asymptotically pivotal. This distribution can be estimated in various ways. We discuss two of them, one based on asymptotic theory and one based on the pairs bootstrap, in the Appendix. For now, we simply let $F\big(\cdot, TT(\beta_0)\big)$ denote the estimated cumulative distribution function (CDF) of $\mathrm{LR}_0(\beta_0)$ conditional on $TT(\beta_0)$, and let c_α denote the $1-\alpha$ quantile of that CDF.

The P value for the hypothesis $\beta = \beta_0$ is $1 - F\big(\mathrm{LR}_0(\beta_0), TT(\beta_0)\big)$, and so the confidence set at nominal confidence level $1-\alpha$ is

$$\big\{\beta_0 \,|\, 1 - F\big(\mathrm{LR}_0(\beta_0), TT(\beta_0)\big) > \alpha\big\}. \tag{14}$$

Using (13), we can replace $\mathrm{LR}_0(\beta_0)$ by $M - TT(\beta_0)$. The inequality inside the braces in (14) can then be rearranged as

$$F\big(M - TT(\beta_0), TT(\beta_0)\big) \le 1 - \alpha. \tag{15}$$

It is shown in the Appendix that, for given M, the function $F(M-c, c)$ decreases monotonically for $0 \le c \le M$. This implies that the equation $F(M-c,c) = 1-\alpha$ has a unique solution $c_\alpha \in [0, M]$ for given α and M, provided that $\alpha > 1 - F(M, 0)$. Because $F(M-c,c)$ is decreasing in c, it follows that the inequality (15) is satisfied for all β_0 such that $TT(\beta_0) \ge c_\alpha$.

The values of β_0 that satisfy the inequality $TT(\beta_0) \ge c_\alpha$ can now be found. By using the Q_{ij} and N_{ij} from (7) in the definition of $TT(\beta_0)$, it can be seen after some algebra that

$$TT(\beta_0) = \frac{n}{\Delta} \frac{A\beta_0^2 - 2B\beta_0 + C}{M_{22}\beta_0^2 - 2M_{12}\beta_0 + M_{11}},$$

where

$$\begin{aligned} A &= P_{22}M_{12}^2 - 2P_{12}M_{12}M_{22} + P_{11}M_{22}^2, \\ B &= P_{22}M_{11}M_{12} - P_{12}M_{11}M_{22} - P_{12}M_{12}^2 + P_{11}M_{12}M_{22}, \text{ and} \\ C &= P_{22}M_{11}^2 - 2P_{12}M_{11}M_{12} + P_{11}M_{12}^2. \end{aligned} \tag{16}$$

The inequality $TT(\beta_0) \ge c_\alpha$ is then equivalent to the quadratic inequality

$$\big(\tfrac{1}{n}M_{22}\Delta c_\alpha - A\big)\beta_0^2 - 2\big(\tfrac{1}{n}M_{12}\Delta c_\alpha - B\big)\beta_0 + \tfrac{1}{n}M_{11}\Delta c_\alpha - C \le 0. \tag{17}$$

Observe that, since $TT(\beta_0)$ is positive, all real values of β_0 must satisfy the inequality $TT(\beta_0) \ge 0$, so that, when $c_\alpha = 0$, the confidence set is the entire real line, $\mathbb{R}$. As is shown in the Appendix, there always exists α small enough that the confidence set

is $\mathbb{R}$. If α is so small that the inequality $\alpha > 1 - F(M, 0)$ is not satisfied, then the confidence set is again $\mathbb{R}$.

If the quadratic equation that sets the left-hand side of (17) to zero has real roots, they are

$$b_\pm = \frac{M_{12}\Delta c_\alpha/n - B}{M_{22}\Delta c_\alpha/n - A} \pm \sqrt{D}, \text{ where} \tag{18}$$

$$D \equiv \frac{(M_{12}\Delta c_\alpha/n - B)^2}{(M_{22}\Delta c_\alpha/n - A)^2} - \frac{M_{11}\Delta c_\alpha/n - C}{M_{22}\Delta c_\alpha/n - A}. \tag{19}$$

If D here is negative, then the left-hand side of (17) is either everywhere positive or everywhere negative. But if it is positive, then (17) cannot be satisfied, which would imply an empty confidence set, contrary to the fact already established that the LIML estimate $\hat{\beta}$ always belongs to the confidence set for any α. This implies that, if $D < 0$, then the coefficient $M_{22}\Delta c_\alpha/n - A$ in (17) must also be negative.

If $D > 0$, the $b_\pm$ are the boundary points of the confidence set. If $M_{22}\Delta c_\alpha/n - A > 0$, the set is the bounded interval $[b_-, b_+]$. If $M_{22}\Delta c_\alpha/n - A < 0$, it is the real line with a hole in it, the hole being the same bounded interval. In the knife-edge case in which $M_{22}\Delta c_\alpha/n - A = 0$, the confidence set is an unbounded interval which may be open either to the left or to the right, depending on the signs of the other coefficients in (17).

We now set out explicitly an algorithm for constructing CLR confidence sets.

1. Compute the six quantities defined in (6) and use them to calculate the quantities A, B, and C defined in (16).
2. Compute $M = TT(\beta_0)$ using (7), (8), (9), (10), and (11).
3. Obtain either the asymptotic or pairs bootstrap critical value c_α using one of the procedures discussed in the Appendix.
4. Evaluate D defined in (19). If $D < 0$, the confidence set is $\mathbb{R}$.
5. If $D > 0$, compute b_- and b_+ using (18). If $M_{22}\Delta c_\alpha/n - A > 0$, the set is the bounded interval $[b_-, b_+]$. Otherwise (ignoring the knife-edge case), it is the real line except for the bounded interval $[b_-, b_+]$.

Note that the CLR confidence set, when it is a bounded interval, is not centered at $\hat{\beta}_{\text{LIML}}$, although, as we have seen, it must always contain $\hat{\beta}_{\text{LIML}}$.

4. RE Bootstrap Confidence Sets

In Davidson and MacKinnon (2008), we proposed the restricted efficient, or RE, bootstrap in the context of hypothesis tests on β in equation (1). In this section, we discuss how the RE bootstrap can also be used to form confidence sets. The simulation results of the next section suggest that confidence sets based on the RE bootstrap generally perform quite well, at least when the instruments are not very

weak. The main disadvantage of these confidence sets is that they are relatively complicated and expensive to compute.

The RE bootstrap has two key features. The bootstrap DGP is conditional on a particular value of β (hence "restricted"), and it uses an efficient estimate of $\boldsymbol{\pi}$ (hence "efficient"). For any specified value β_0, we can run regression (1) to obtain parameter estimates $\tilde{\boldsymbol{\gamma}}$ and residuals $\tilde{\boldsymbol{u}}_1$. The latter may be rescaled by multiplying them by a factor of $(n/(n-k))^{1/2}$. We then run the regression

$$\boldsymbol{y}_2 = \boldsymbol{W}\boldsymbol{\pi} + \delta\tilde{\boldsymbol{u}}_1 + \text{residuals}. \tag{20}$$

This yields parameter estimates $\tilde{\boldsymbol{\pi}}$ and adjusted residuals $\tilde{\boldsymbol{u}}_2 \equiv \boldsymbol{y}_2 - \boldsymbol{W}\tilde{\boldsymbol{\pi}}$. The latter should be rescaled by multiplying them by a factor of $(n/(n-l))^{1/2}$. It can be shown that $\tilde{\boldsymbol{\pi}}$ is asymptotically equivalent to the estimate one would obtain by using FIML or 3SLS. This estimate was used by Kleibergen (2002) in a different context.

Generating a bootstrap sample using the RE bootstrap is quite simple. We form two vectors of bootstrap disturbances, $\boldsymbol{u}_1^*$ and $\boldsymbol{u}_2^*$, with elements u_{i1}^* and u_{i2}^* for $i = 1, \ldots, n$, resampled from the pairs of rescaled residuals. We then set

$$\begin{aligned} \boldsymbol{y}_2^* &= \boldsymbol{W}\tilde{\boldsymbol{\pi}} + \boldsymbol{u}_2^*, \text{ and} \\ \boldsymbol{y}_1^* &= \beta_0\boldsymbol{y}_2^* + \boldsymbol{Z}\tilde{\boldsymbol{\gamma}} + \boldsymbol{u}_1^*. \end{aligned} \tag{21}$$

If we generate B bootstrap samples, we can compute an equal-tail bootstrap P value for the hypothesis that $\beta = \beta_0$. It is simply

$$\hat{p}^*(\beta_0) = \frac{2}{B}\min\left(\sum_{j=1}^{B} \mathrm{I}(\tau_j^* < \hat{\tau}),\ \sum_{j=1}^{B} \mathrm{I}(\tau_j^* \geq \hat{\tau})\right), \tag{22}$$

where $\mathrm{I}(\cdot)$ is the indicator function, $\hat{\tau} = (\hat{\beta} - \beta_0)/s(\hat{\beta})$, and $\tau_j^* = (\hat{\beta}_j^* - \beta_0)/s(\hat{\beta}_j^*)$. Here $\hat{\beta}$ may denote either $\hat{\beta}_{\mathrm{IV}}$ or $\hat{\beta}_{\mathrm{LIML}}$, and $\hat{\beta}_j^*$ then denotes the corresponding estimate for the j^{th} bootstrap sample. It is important to calculate the standard errors $s(\hat{\beta})$ and $s(\hat{\beta}_j^*)$ in the same way. By using the equal-tail P value (22), we do not impose symmetry on the distribution of τ.

Using the RE bootstrap to obtain a confidence set is a bit complicated. Consider the upper limit, $\hat{\beta}_u$. Start with an initial estimate, say $\hat{\beta}_u^1$ (one obvious candidate is the upper limit of the asymptotic confidence interval) and compute $\hat{p}^*(\hat{\beta}_u^1)$ using equation (22). If $\hat{p}^*(\hat{\beta}_u^1) > \alpha$, then $\hat{\beta}_u^1$ is too small; if $\hat{p}^*(\hat{\beta}_u^1) < \alpha$, then it is too large. Try another candidate, say $\hat{\beta}_u^2$, which must be larger than $\hat{\beta}_u^1$ in the former case and smaller in the latter case. Calculate $\hat{p}^*(\hat{\beta}_u^2)$ and repeat if necessary. The way in which $\hat{\beta}_u^2$ is chosen may have a significant impact on computational cost, but it should have no effect on the properties of the RE bootstrap confidence set.

If, after m tries, we have found $\hat{\beta}_u^{m-1}$ and $\hat{\beta}_u^m$ such that $\hat{p}^*(\hat{\beta}_u^{m-1}) - \alpha$ and $\hat{p}^*(\hat{\beta}_u^m) - \alpha$ have opposite signs, then $\hat{\beta}_u$ must lie between them. At this point, various numerical

methods can be used to find it. Since $\hat{p}^*(\beta_0)$ is not differentiable, we must use a method that does not need derivatives. In this paper, we use bisection, which is easy to program and reasonably fast. Note that exactly the same set of random numbers must be used for all the bootstrap samples. Otherwise, the value of $\hat{p}^*(\beta_0)$ would be different each time we evaluated it.

The procedure for finding the lower limit, $\hat{\beta}_l$, is essentially the same as the one for finding the upper limit, with obvious changes in sign at various points.

In the above description of the algorithm, we have implicitly assumed that, if β_0 is sufficiently large or sufficiently small, $\hat{p}^*(\beta_0)$ must be less than α. However, that is not always true. The confidence set has no upper bound if $p^*(\beta_0) > \alpha$ as β_0 tends to plus infinity, and it has no lower bound if $p^*(\beta_0) > \alpha$ as β_0 tends to minus infinity. In practice, we may reasonably conclude that the confidence set is unbounded from above (below) if $p^*(\beta_0) > \alpha$ for a very large positive (negative) value of β_0.

Like CLR confidence sets, unbounded RE bootstrap confidence sets may contain holes. In fact, simulations suggest that they frequently contain a hole when they are unbounded. It is therefore important to check for holes and for unboundedness even if the procedure described above has apparently located both $\hat{\beta}_u$ and $\hat{\beta}_l$. If there are values of β_0 greater than $\hat{\beta}_u$ or less than $\hat{\beta}_l$ for which $p^*(\beta_0) > \alpha$, it is easy enough to locate the other end of the hole. However, we do not recommend using unbounded confidence sets to make inferences. The fact that a confidence set is unbounded strongly suggests that the instruments are so weak as to make reliable inference impossible.

The fact that RE bootstrap confidence sets may be unbounded (and in fact often are unbounded when the instruments are very weak) is actually a desirable feature; see Gleser and Hwang (1987) and Dufour (1997). We noted this fact earlier in the context of CLR confidence sets. Because RE bootstrap confidence sets can be unbounded, it is possible for them to have very good coverage.

Unless heteroskedasticity is clearly absent, it is generally wise to use confidence sets that are robust to it. One advantage of using confidence sets based on t statistics is that it is very easy to do so. We simply replace the ordinary t statistic with one based on a heteroskedasticity-consistent standard error and employ a slightly modified version of the RE bootstrap.

The wild restricted efficient, or WRE, bootstrap was proposed by Davidson and MacKinnon (2010). It is very similar to the RE bootstrap, except that the i^{th} pair of rescaled residuals is always associated with the i^{th} observation. To generate the bootstrap disturbances, we simply multiply each pair of rescaled residuals by a random variable v_i^* with mean zero and variance one. See Davidson and Flachaire (2008) for more about the wild bootstrap. In samples of reasonable size (more than a few hundred observations) with heteroskedastic disturbances, this should work just about as well as using ordinary standard errors and the RE bootstrap when the disturbances are actually homoskedastic.

5. Simulation Evidence

Following Davidson and MacKinnon (2008), we use the DGP:

$$\begin{aligned} \boldsymbol{y}_1 &= \beta \boldsymbol{y}_2 + \boldsymbol{u}_1, \\ \boldsymbol{y}_2 &= a\boldsymbol{w} + \boldsymbol{u}_2, \end{aligned} \tag{23}$$

where $\boldsymbol{w} \in \mathcal{S}(\boldsymbol{W})$ is an n–vector with $\|\boldsymbol{w}\|^2 = 1$, and

$$\begin{aligned} \boldsymbol{u}_1 &= r\boldsymbol{v}_1 + \rho\boldsymbol{v}_2, \\ \boldsymbol{u}_2 &= \boldsymbol{v}_2, \end{aligned} \qquad \begin{bmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \end{bmatrix} \sim \mathrm{N}(\boldsymbol{0}, \mathbf{I}), \quad r^2 + \rho^2 = 1. \tag{24}$$

It may seem curious that there is just a single instrument $\boldsymbol{w}$ in the DGP when there are l of them in equation (2). But the only property of $\boldsymbol{W}$ that matters is $\mathcal{S}(\boldsymbol{W})$, the subspace spanned by the columns of $\boldsymbol{W}$. In effect, we have performed a linear transformation on $\boldsymbol{W}$ so that all of the explanatory power comes from the vector $\boldsymbol{w}$ and the other columns of $\boldsymbol{W}$ are simply noise. Of course, such a transformation has no effect on $\mathcal{S}(\boldsymbol{W})$.

The instrument vector $\boldsymbol{w}$ is normalized to have squared length unity; that is, $\boldsymbol{w}^\top\boldsymbol{w} = 1$. By employing this normalization, we are implicitly using weak-instrument asymptotics; see Staiger and Stock (1997). The strength of the instruments is measured by the parameter a, the square of which is the scalar concentration parameter; see Phillips (1983, p. 470) and Stock, Wright, and Yogo (2002). Because we are only concerned with confidence sets, the error variances have all been normalized to unity, which is something we could not do if we were concerned with bias.

The first three figures each contain six panels, two for each of the IV Wald, IV LIML, and CLR confidence sets. In all cases, the left-hand panel of each pair shows coverage for asymptotic 95% confidence sets (which are based on the standard normal distribution for the two Wald intervals), and the right-hand panel shows coverage for 95% confidence sets based on the pairs bootstrap. Asymptotic results are based on 500,000 replications, and bootstrap results are based on 100,000 replications, each with $B = 999$ bootstrap samples.

Figure 1 shows the effect of varying the number of instruments that are not also regressors in the structural equation, that is, $l - k$, for six values of a^2. The six values are 4, 8, 16, 32, 64, and 128, and $l - k$ varies from 1 to 18. The sample size is fairly large ($n = 400$), and the correlation between the structural and reduced-form errors is quite high ($\rho = 0.8$).

For the asymptotic IV Wald intervals, there is generally severe undercoverage unless $l - k$ is small and a^2 is large. The pairs bootstrap generally helps somewhat, except when $l - k$ is small. For the bootstrap intervals, undercoverage is moderate when $a^2 \geq 64$ and $l - k \leq 10$, but it is still severe in most cases.

The asymptotic LIML Wald intervals always work better than the corresponding asymptotic IV ones. Undercoverage is nonexistent or quite moderate when $a^2 \geq 64$

for all values of $l - k$. However, the pairs bootstrap actually makes undercoverage worse, especially for small values of a^2.

The asymptotic CLR confidence sets perform extremely well. They always undercover, but only very slightly when $l - k$ is small. The undercoverage gradually increases as $l - k$ increases, especially when a^2 is small, but it is never very great. In contrast, the pairs bootstrap CLR confidence sets almost always overcover, and they do so severely when a^2 is small and $l - k$ is large.

Figure 2 shows the effect of changing the sample size for $l - k = 10$, $\rho = 0.8$, and the same six values of a^2. The sample sizes are 50, 70, 100, 141, 200, 282, 400, 565, 800, 1131, and 1600; each of these is approximately $\sqrt{2}$ times its predecessor. The performance of all the Wald intervals is strikingly insensitive to sample size. There tend to be slight improvements in coverage as n increases, which is most noticeable for $a^2 = 128$.

In contrast, the performance of the CLR confidence sets depends greatly on the sample size. The undercoverage of the asymptotic CLR confidence sets diminishes rapidly as n increases. The overcoverage of the pairs bootstrap CLR intervals also diminishes, but less rapidly, especially for the smaller values of a^2.

Figure 3 shows the effect of changing ρ. For the asymptotic results, there are 100 values between 0.00 and 0.99 increasing by 0.01. For the bootstrap results, there are 34 values between 0.00 and 0.99 increasing by 0.03. The sample size is 50, and $l - k = 10$.

Coverage of all the confidence sets depends strongly on ρ, except sometimes when a^2 is large. This is most true for the asymptotic IV Wald intervals, which actually overcover for both a^2 and ρ small, even though they undercover very severely for a^2 small and ρ large. As in Figure 1, the pairs bootstrap generally improves coverage for IV Wald intervals (but not when ρ is small). However, except when a^2 is large, it actually causes LIML Wald intervals to undercover more severely.

The CLR confidence sets are only moderately sensitive to ρ. They do not perform particularly well in Figure 3, because the sample size is only 50. Based on the results in Figure 2, we can be confident that the undercoverage of asymptotic CLR intervals would be very much less severe if n were substantially larger.

Figure 4 shows the coverage of RE bootstrap confidence sets based on both IV and LIML t statistics for the same case as Figure 1, that is, $\rho = 0.80$, $n = 400$, and $l - k$ varying between 1 and 18. Note the scale of the vertical axis. Although coverage is certainly not perfect, it is vastly better than for the asymptotic and pairs bootstrap Wald intervals. For larger values of $l - k$ and a^2, it is even better than for the asymptotic CLR confidence sets.

Figure 5 shows the coverage of RE bootstrap confidence sets for the same case as Figure 2, that is, $\rho = 0.80$, $l - k = 10$, and n varying between 50 and 1600. Once again, coverage is vastly better than for the asymptotic and pairs bootstrap Wald

intervals. It is also better than the coverage of the CLR confidence sets for most sample sizes, but not for the largest sample sizes when a^2 is small.

The results for the IV and LIML cases in Figure 5 are often quite different, and there are a few results that are hard to explain. The LIML confidence sets all work essentially perfectly for $n \geq 200$ and $a^2 \geq 32$. However, it is only for $a^2 \geq 128$ that we can make a similar statement for the IV confidence sets. Not coincidentally, all of the LIML confidence sets for $n \geq 200$ are bounded for $a^2 \geq 64$, and nearly all are bounded when $a^2 = 32$, while a significant fraction of the IV confidence sets are unbounded even when $a^2 = 64$.

Figure 6 shows the coverage of RE bootstrap confidence sets for the same case as Figure 3, that is, $n = 50$, $l - k = 10$, and ρ varying from 0.00 to 0.99 by 0.03. Once again, coverage is very much better in most cases than it was in Figure 3. It generally deteriorates as ρ increases, especially for very large values of ρ in the LIML case when $a^2 \leq 16$.

Figures 7 and 8 report results from a different set of experiments in which $n = 400$ and $l-k = 2$. Thus the sample size is fairly large, and there is only one overidentifying restriction. This is a situation which may be typical of quite a few applied studies, and in which we would expect all of the better methods to work well. We do not report results for coverage, because they do not vary a lot with ρ and are therefore similar to the results for $l - k = 2$ in Figures 1 and 4. The figures show results only for $a^2 = 8$ and $a^2 = 16$. For larger values of a^2, the asymptotic CLR and RE bootstrap LIML Wald intervals are bounded almost all the time. That is also the case for the RE bootstrap IV Wald intervals for $a^2 \geq 64$.

Figure 7 shows the fraction of confidence sets that are bounded intervals. This fraction is highest for the CLR intervals and lowest for the RE bootstrap IV Wald intervals. In the case of the latter, it drops sharply as ρ increases. Figure 8 shows the median length of bounded confidence intervals. The CLR intervals are, on average, the longest when $a^2 = 8$, probably because there are quite a few cases in which the CLR interval is bounded and one or both of the RE bootstrap ones are not. When $a^2 = 16$, the CLR intervals continue to be longer than the bootstrap LIML ones, but they are a little bit shorter than the bootstrap IV ones for small values of ρ. In both cases, the median length of the bootstrap IV intervals drops sharply as ρ increases, presumably because the fraction of confidence sets that are bounded also drops sharply.

The figures do not show results for the two largest values of a^2 (i.e., reasonably strong instruments), because they would be very hard to read. In these cases, all three confidence sets are always bounded (almost always for IV when $a^2 = 64$), and their lengths are very similar, with CLR tending to be very slightly shorter.

Because the performance of the various confidence sets depends on so many aspects of the experimental design (n, a^2, $l - k$, and ρ), it is difficult to draw definitive conclusions. Nevertheless, we tentatively draw several results.

- Asymptotic CLR confidence sets seem to perform remarkably well whenever the sample size is sufficiently large, even when the instruments are very weak. However, there can be substantial undercoverage when the sample size is small. In contrast, pairs bootstrap CLR confidence sets always overcover, often severely, even when the sample size is very large.
- RE bootstrap confidence sets based on IV and LIML t statistics perform very much better than either asymptotic or pairs bootstrap confidence intervals based on the same test statistics, even when the sample size is small.
- RE bootstrap confidence sets based on LIML t statistics are generally preferable to ones based on IV t statistics, even though their coverage may be either better or worse. They more frequently consist of a single, bounded interval, and they tend to be shorter whenever the instruments are strong enough that all or almost all the confidence sets of both types are bounded intervals. However, when the instruments are strong enough for this to be the case, asymptotic CLR intervals seem to be slightly shorter than RE bootstrap LIML ones.

6. Conclusion

We have proposed a new bootstrap procedure for constructing confidence sets for the coefficient of the single right-hand-side endogenous variable in a linear equation with weak instruments. This procedure is based on the RE bootstrap that was proposed in the context of hypothesis testing in Davidson and MacKinnon (2008). A very similar procedure based on the WRE bootstrap of Davidson and MacKinnon (2010) can be used when there may be heteroskedasticity of unknown form. We have also provided a new derivation of, and computational procedure for, the asymptotic CLR confidence interval proposed by Mikusheva (2010), along with a pairs bootstrap variant.

Even though the new RE bootstrap procedure is based on t statistics, it generally produces quite reliable confidence sets. These have far better coverage than asymptotic and pairs bootstrap intervals based on the same test statistics. For small sample sizes, they are often more reliable than asymptotic CLR confidence sets. For large sample sizes, however, the latter seem to be slightly preferable, especially when the instruments are very weak, and the CLR intervals are certainly much less computationally intensive.

One important advantage of the RE bootstrap procedure is that it can easily be modified to handle heteroskedasticity of unknown form. In principle, it can also deal with cases in which there are two or more endogenous variables on the right-hand side of a structural equation. These are both subjects for further research, as is the possibility of using the RE bootstrap to form CLR confidence sets.

Appendix: Asymptotic and Bootstrap CLR Critical Values

The asymptotic approximation

The distribution of the CLR statistic $\mathrm{LR}_0(\beta)$ evaluated at $\beta = \beta_0$, with β_0 the true value of the parameter, does not depend on β_0. Similarly, the distribution of $\mathrm{LR}_0(\beta_0)$ conditional on $TT(\beta_0)$ does not depend on β_0. Thus we can without loss of generality set $\beta = 0$ in what follows, and drop the explicit dependence of LR_0, SS, ST, and TT on β_0. We denote the CDF of LR_0 conditional on TT by $F(\cdot, TT)$.

We begin with the asymptotic approximation to $F(\cdot, TT)$. It is shown in Moreira (2003) and in Davidson and MacKinnon (2008) that, asymptotically, the random variables $Z \equiv ST/\sqrt{TT}$ and $Y \equiv SS - ST^2/TT$ are independent conditionally on TT, with distributions $\mathrm{N}(0,1)$ and χ^2_{l-k-1}, respectively. It is easy to see that

$$\begin{aligned} 2\,\mathrm{LR}_0 &= SS - TT + \sqrt{(SS - TT)^2 + 4ST^2} \\ &= Y + Z^2 - TT + \sqrt{(Y + Z^2 - TT)^2 + 4TT\,Z^2}. \end{aligned} \tag{A1}$$

Thus

$$\begin{aligned} F(x, TT) &= \mathrm{E}\big[\mathrm{I}(\mathrm{LR}_0 \le x) \,|\, TT\big] \\ &= \mathrm{E}\big[\mathrm{I}(Y + Z^2 - TT + \sqrt{(Y + Z^2 - TT)^2 + 4TT\,Z^2} \le 2x) \,|\, TT\big]. \end{aligned}$$

The inequality in the indicator function above can be rewritten as

$$\sqrt{(Y + Z^2 - TT)^2 + 4TT\,Z^2} \le 2x - (Y + Z^2 - TT).$$

Since the left-hand side is the positive square root, this is equivalent to

$$(Y + Z^2 - TT)^2 + 4TT\,Z^2 \le (Y + Z^2 - TT)^2 - 4x(Y + Z^2 - TT) + 4x^2,$$

which, since the first term on each side of the inequality is the same, implies that

$$Y \le x + TT - Z^2(1 + TT/x) = (x + TT)(1 - Z^2/x).$$

We can now make use of the asymptotic conditional distributions of Y and Z to compute the asymptotic approximation to $F(x, TT)$. Since asymptotically $Y \sim \chi^2_{l-k-1}$,

$$\begin{aligned} F(x, TT) &= \mathrm{E}\big[\mathrm{E}\big(\mathrm{I}(Y \le (x + TT)(1 - Z^2/x) \,|\, Z\big) \,|\, TT\big] \\ &\approx \mathrm{E}\big[F_{\chi^2_{l-k-1}}\big((x + TT)(1 - Z^2/x)\big) \,|\, TT\big], \end{aligned} \tag{A2}$$

where $F_{\chi^2_{l-k-1}}$ is the CDF of χ^2_{l-k-1}. The argument of this CDF is negative if $Z^2 > x$, and so its value is zero. Thus, when we approximate (A2) using the asymptotic distribution $Z \sim \mathrm{N}(0,1)$, the result is

$$\begin{aligned} F_{\mathrm{as}}(x, TT) &\equiv \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{x}}^{\sqrt{x}} F_{\chi^2_{l-k-1}}\big((x+TT)(1-z^2/x)\big) \mathrm{e}^{-z^2/2}\,\mathrm{d}z \\ &= \sqrt{\frac{2}{\pi}} \int_{0}^{\sqrt{x}} F_{\chi^2_{l-k-1}}\big((x+TT)(1-z^2/x)\big) \mathrm{e}^{-z^2/2}\,\mathrm{d}z. \end{aligned} \tag{A3}$$

It is not hard to evaluate this for given arguments x and TT by numerical integration. Mikusheva (2010) and Andrews, Moreira, and Stock (2007) use the following expression for this approximation:

$$F(x, TT) \approx 2K_4 \int_0^1 F_{\chi^2_l}\big((x+TT)/(1+TT\,z^2/x)\big)(1-z^2)^{(l-k-3)/2}\,\mathrm{d}z, \tag{A4}$$

where $K_4 = 1/\mathrm{B}(1/2, (l-k-1)/2)$, B being the beta function. Expressions (A3) and (A4) are equal, although derived in different ways.

Recall from (14) that the critical value c_α used to construct the CLR confidence set solves the equation $F(M-c, c) = 1-\alpha$, provided a solution exists and is unique. We now show that, for given M, the function $F_{\mathrm{as}}(M-c, c)$ decreases monotonically for $0 \le c \le M$. From (A3), we have

$$F_{\mathrm{as}}(M-c, c) = \sqrt{\frac{2}{\pi}} \int_0^{\sqrt{M-c}} F_{\chi^2_{l-k-1}}\big(M(1-z^2/(M-c))\big)\,\mathrm{e}^{-z^2/2}\,\mathrm{d}z, \tag{A5}$$

from which it is clear that, for $c = M$, $F_{\mathrm{as}}(M-c, c) = F_{\mathrm{as}}(0, M) = 0$. The integrand in (A5) when evaluated at the upper limit $z = \sqrt{M-c}$ is zero, and so the derivative of $F_{\mathrm{as}}(M-c, c)$ with respect to c is

$$-\sqrt{\frac{2}{\pi}} \int_0^{\sqrt{M-c}} f_{\chi^2_{l-k-1}}\big(M(1-z^2/(M-c))\big) \frac{Mz^2}{(M-c)^2}\,\mathrm{e}^{-z^2/2}\,\mathrm{d}z, \tag{A6}$$

where $f_{\chi^2_{l-k-1}}$ is the density of χ^2_{l-k-1}. It is obvious that this derivative is negative everywhere for $0 < c \le M$.

For $c = 0$, (A5) becomes

$$F_{\mathrm{as}}(M, 0) = \sqrt{\frac{2}{\pi}} \int_0^{\sqrt{M}} F_{\chi^2_{l-k-1}}(M-z^2) \mathrm{e}^{-z^2/2}\,\mathrm{d}z.$$

With the change of integration variable $z^2 = y$, this is

$$\frac{1}{\sqrt{2\pi}} \int_0^M F_{\chi^2_{l-k-1}}(M-y) \frac{\mathrm{e}^{-y/2}}{\sqrt{y}}\,\mathrm{d}y = \int_0^M F_{\chi^2_{l-k-1}} f_{\chi^2_1}(y)\,\mathrm{d}y,$$

where $f_{\chi^2_1}$ is the density of χ^2_1. The last expression is thus a convolution, expressing the CDF of the distribution of the sum of a χ^2_{l-k-1} variable and an independent χ^2_1 variable, that is, of a χ^2_{l-k} variable, evaluated at M. Thus $F_{\mathrm{as}}(M, 0) = F_{\chi^2_{l-k}}(M)$.

These properties make it clear that the equation $F_{\mathrm{as}}(M - c, c) = 1 - \alpha$ has a unique solution $c_\alpha \in [0, M]$ for given α and M, provided that

$$\alpha > 1 - F_{\mathrm{as}}(M, 0) = 1 - F_{\chi^2_{l-k}}(M). \tag{A7}$$

The confidence set includes all β_0 for which $TT(\beta_0) \geq c_\alpha$. We saw previously that, since $TT(\beta_0)$ is non-negative, the confidence set is the whole real line when $c_\alpha = 0$, which is the case when $\alpha = 1 - F_{\chi^2_{l-k}}(M)$. Since $F_{\chi^2_{l-k}}(M) < 1$ for finite M, there always exists α small enough that the confidence set is $\mathbb{R}$. If α is so small that (A7) is not satisfied, then *a fortiori* the confidence set is again $\mathbb{R}$.

Solving for the critical value c_α

The derivative (A6) of $F_{\mathrm{as}}(M - c, c)$ with respect to c can be expressed in terms of elementary functions and the gamma function only, since, for any positive d, the density of χ^2_d is

$$f_{\chi^2_d}(x) = \frac{1}{2^{(d/2}\Gamma(d/2)} x^{d/2-1} \mathrm{e}^{-x/2},$$

where Γ is the gamma function. Therefore, the derivative (A6) is

$$-\frac{\mathrm{e}^{-M/2} M^{(l-k-1)/2}}{2^{(l-k)/2-1}\sqrt{\pi(M-c)}\Gamma((l-k-1)/2)} \int_0^1 (1-y^2)^{(l-k-3)/2} y^2 \mathrm{e}^{cy^2/2}\, \mathrm{d}y.$$

This expression, although messy in appearance, is readily evaluated numerically.

The ability to do so means that solving the equation $F_{\mathrm{as}}(M - c, c) = 1 - \alpha$ can be done by Newton's method, as well as by more basic methods such as bisection. For any such method, it is good to have a starting point reasonably close to the actual solution. The graph of the function $F_{\mathrm{as}}(M - c, c)$, for large M at least, resembles an inverted 'L', with the value of the function close to $F_{\chi^2_{l-k}}(M)$ for all values of c until c is close to M, at which point the graph suddenly curves almost vertically downward to 0 as $c \to M$. If we change the integration variable in (A5) by the formula $y = z\sqrt{M/(M-c)}$, we see that

$$F_{\mathrm{as}}(M - c, c) = \sqrt{\frac{2(M-c)}{\pi M}} \int_0^{\sqrt{M}} F_{\chi^2_{l-k-1}}(M - y^2) \exp\Big(-\frac{y^2(M-c)}{2M}\Big)\, \mathrm{d}y, \tag{A8}$$

which depends on c only through the difference $M - c$. We expect that difference to be small relative to M, on the basis of the appearance of the graph.

In order to find c_α, we need to solve equation (A8). For the purpose of an approximation to the solution, let us expand the exponential in the integrand, retaining only the first two terms of the expansion. This gives the approximate expression

$$F_{\text{as}}(M-c,c) \approx \sqrt{\frac{2(M-c)}{\pi M}} \int_0^{\sqrt{M}} F_{\chi^2_{l-k-1}}(M-y^2)\Big(1-\frac{y^2(M-c)}{2M}\Big)\,\mathrm{d}y.$$

Now make the definitions

$$K_1 = \sqrt{\frac{2}{\pi M}} \int_0^{\sqrt{M}} F_{\chi^2_{l-k-1}}(M-y^2)\,\mathrm{d}y$$

and

$$K_2 = \sqrt{\frac{1}{2\pi M^3}} \int_0^{\sqrt{M}} y^2 F_{\chi^2_{l-k-1}}(M-y^2)\,\mathrm{d}y.$$

It is not hard to evaluate K_1 and K_2 for given M by numerical integration. The equation we wish to solve for c_α is approximated by

$$1-\alpha = K_1\sqrt{M-c} - K_2(M-c)^{3/2}. \tag{A9}$$

A first approximation to the solution of this equation is just $c = M - ((1-\alpha)/K_1)^2$, where we retain only the first term on the right-hand side of (1). A better approximation is obtained by retaining both terms and using the first approximate solution in the second term. This gives

$$c \approx M - \frac{(1-\alpha)^2}{K_1^2}\Big(1+\frac{K_2(1-\alpha)^2}{K_1^3}\Big)^2. \tag{A10}$$

Numerical experiments show that this is an excellent approximation, starting from which Newton's method usually converges in fewer than 4 or 5 iterations.

In the special case in which $l-k=1$, there is no need to use an iterative procedure to find c_α. In this case, $Y=0$, which by (A1) implies that LR_0 is equal to Z^2 independently of TT. Thus $F_{\text{as}}(x,c) = F_{\chi^2_1}(x)$, and so the solution to $F_{\text{as}}(M-c,c) = 1-\alpha$ is just $c = M - F^{-1}_{\chi^2_1}(1-\alpha)$.

The bootstrap approximation

The test that Davidson and MacKinnon (2008) call the CLRb test is a bootstrap test in which the conditional distribution of LR_0 is approximated by generating bootstrap statistics of the form

$$\text{LR}_0^* = \frac{1}{2}\Big(SS^* - TT + \sqrt{(SS^*-TT)^2 + 4TT(ST^*)^2/TT^*}\Big) \tag{A11}$$

conditional on TT from the observed data. Here SS^*, ST^*, and TT^* are calculated using (8), (9), and (10) from starred versions of the six quantities defined in (6),

computed using data generated by a bootstrap DGP that is intended to approximate the true DGP for the model specified by (1) and (2), with $\beta = 0$. The conditional CDF $F(x, TT)$ is then approximated by

$$F_{\text{bs}}(x, TT) \equiv \frac{1}{B} \sum_{j=1}^{B} \text{I}\big((\text{LR}_0^*)_j \leq x\big), \tag{A12}$$

and the bootstrap P value is just $1 - F_{\text{bs}}(\text{LR}_0, TT)$.

By inverting this procedure, we can obtain a bootstrap version of the critical value c_α needed for a CLR confidence interval. The simplest approach, which was suggested by Moreira, Porter, and Suarez (2005), is to use the pairs bootstrap DGP described in Section 2. First, each bootstrap sample is used to calculate starred versions of the six quantities defined in (6), which are then used to calculate the quantities Q_{ij}^* and N_{ij}^* for $i = 1, 2$ using (7) with $\beta_0 = \hat{\beta}_{\text{LIML}}$. These in turn are used in (8), (9), and (10) to calculate SS^*, ST^*, and TT^*.

In order to invert the CLRb test, we have to solve the equation $F_{\text{bs}}(M - c, c) = 1 - \alpha$, which can be written more explicitly as

$$\frac{1}{B} \sum_{j=1}^{B} \text{I}\big((\text{LR}_0^*(c))_j \leq M - c\big) = 1 - \alpha. \tag{A13}$$

Here $\text{LR}_0^*(c)$ is computed using formula (A11) with TT replaced by c.

Solving equation (A13) may require computing $\text{LR}_0^*(c)$ for quite a few values of c. Because the sum in (A13) is a discontinuous function of c for finite B, Newton's method is not an appropriate way to solve that equation. However, the approximation (A10) should still provide an excellent starting point for any method that does not use derivatives, such as bisection.

References

Anderson, T. W., and H. Rubin (1949), "Estimation of the parameters of a single equation in a complete system of stochastic equations," *Annals of Mathematical Statistics*, 20, 46–63.

Andrews, D. W. K., M. Moreira, and J. Stock (2007),"Performance of conditional Wald tests in IV regression with weak instruments," *Journal of Econometrics*, 139, 116–132.

Davidson, R., and E. Flachaire (2008), "The wild bootstrap, tamed at last," *Journal of Econometrics*, 146, 162–169.

Davidson, R. and J. G. MacKinnon (2004), *Econometric Theory and Methods*, New York, Oxford University Press.

Davidson, R. and J. G. MacKinnon (2008), "Bootstrap inference in a linear equation estimated by instrumental variables," *Econometrics Journal*, 11, 443–477.

Davidson, R. and J. G. MacKinnon (2010), "Wild bootstrap tests for IV regression," *Journal of Business and Economic Statistics*, 28, 128–144.

Davidson, R. and J. G. MacKinnon (2011), "Confidence sets based on inverting Anderson-Rubin tests," QED Working Paper 1257.

Davison, A. C., and D. V. Hinkley (1997), *Bootstrap Methods and Their Application*, Cambridge, Cambridge University Press.

Dufour, J.-M. (1997), "Some impossibility theorems in econometrics with applications to structural and dynamic models," *Econometrica*, 65, 1365–1388.

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

Gleser, L., and J. Hwang (1987). "The non-existence of 100(1-α)% confidence sets of finite expected diameter in errors-in-variables and related models," *Annals of Statistics*, 15, 1351–1362.

Kleibergen, F. (2002). "Pivotal statistics for testing structural parameters in instrumental variables regression," *Econometrica*, 70, 1781–1803.

Mikusheva, A. (2010). "Robust confidence sets in the presence of weak instruments," *Journal of Econometrics*, 157, 236–247.

Moreira, M. J. (2003). "A conditional likelihood ratio test for structural models," *Econometrica*, 71, 1027–1048.

Moreira, M. J., J. R. Porter and G. A. Suarez (2005). "Bootstrap and higher-order expansion validity when instruments may be weak," NBER Working Paper No. 302, revised.

Phillips, P. C. B. (1983), "Exact small sample theory in the simultaneous equations model," Chp. 8 in Z. Griliches and M. D. Intriligator (eds.), *Handbook of Econometrics*, Vol. 1, pp. 449–516. Amsterdam: North Holland.

Staiger, D., and J. H. Stock (1997), "Instrumental variables regression with weak instruments," *Econometrica*, 65, 557–586.

Stock, J. H., J. H. Wright, and M. Yogo (2002), "A survey of weak instruments and weak identification in generalized method of moments," *Journal of Business and Economic Statistics*, 20, 518–29.

Zivot, E., R. Startz, and C. R. Nelson (1998), "Valid confidence intervals and inference in the presence of weak instruments," *International Economic Review*, 39, 1119–1144.

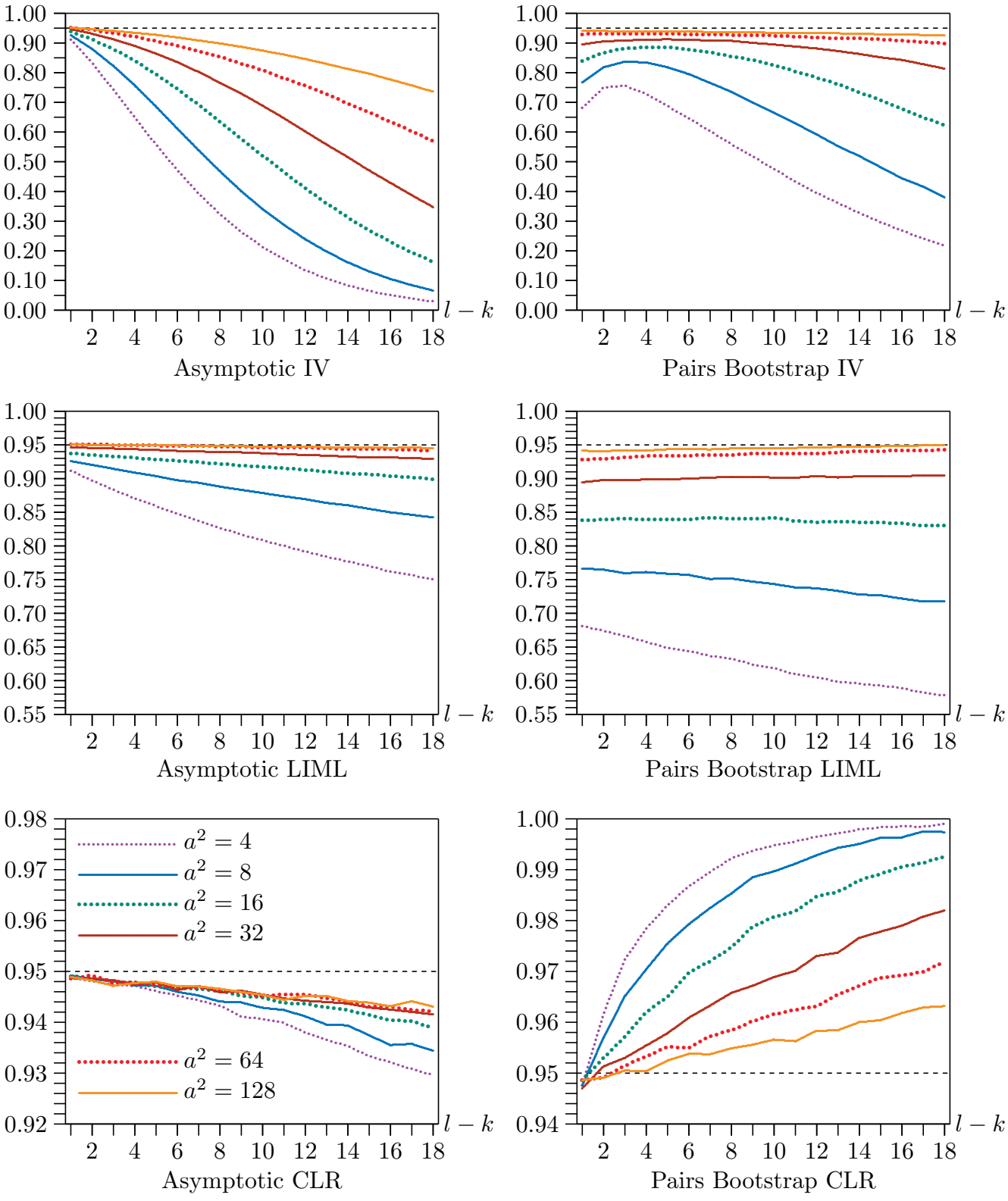

Figure 1. Coverage of confidence sets for $\rho = 0.80$ and $n = 400$

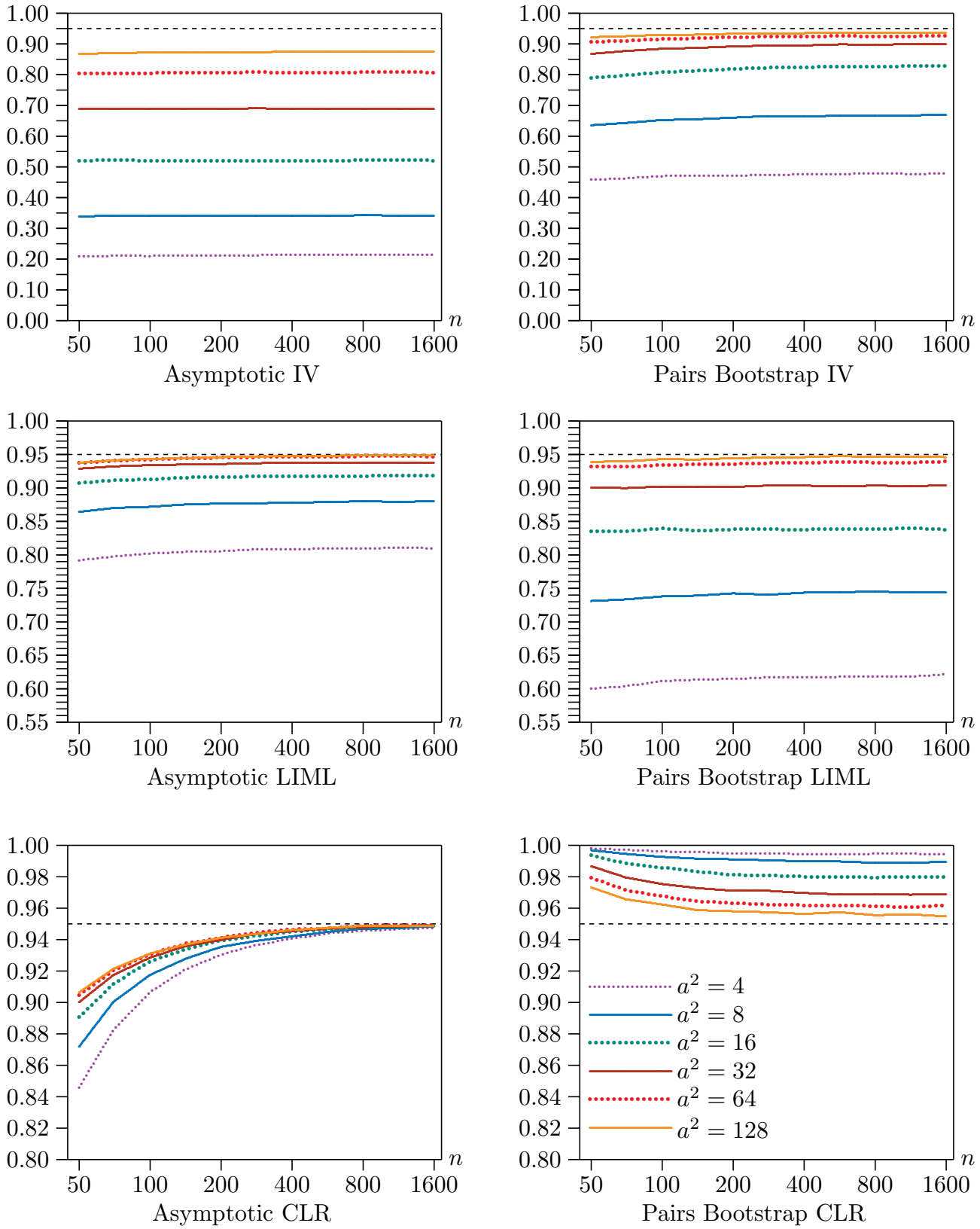

Figure 2. Coverage of confidence sets for $\rho = 0.80$ and $l - k = 10$

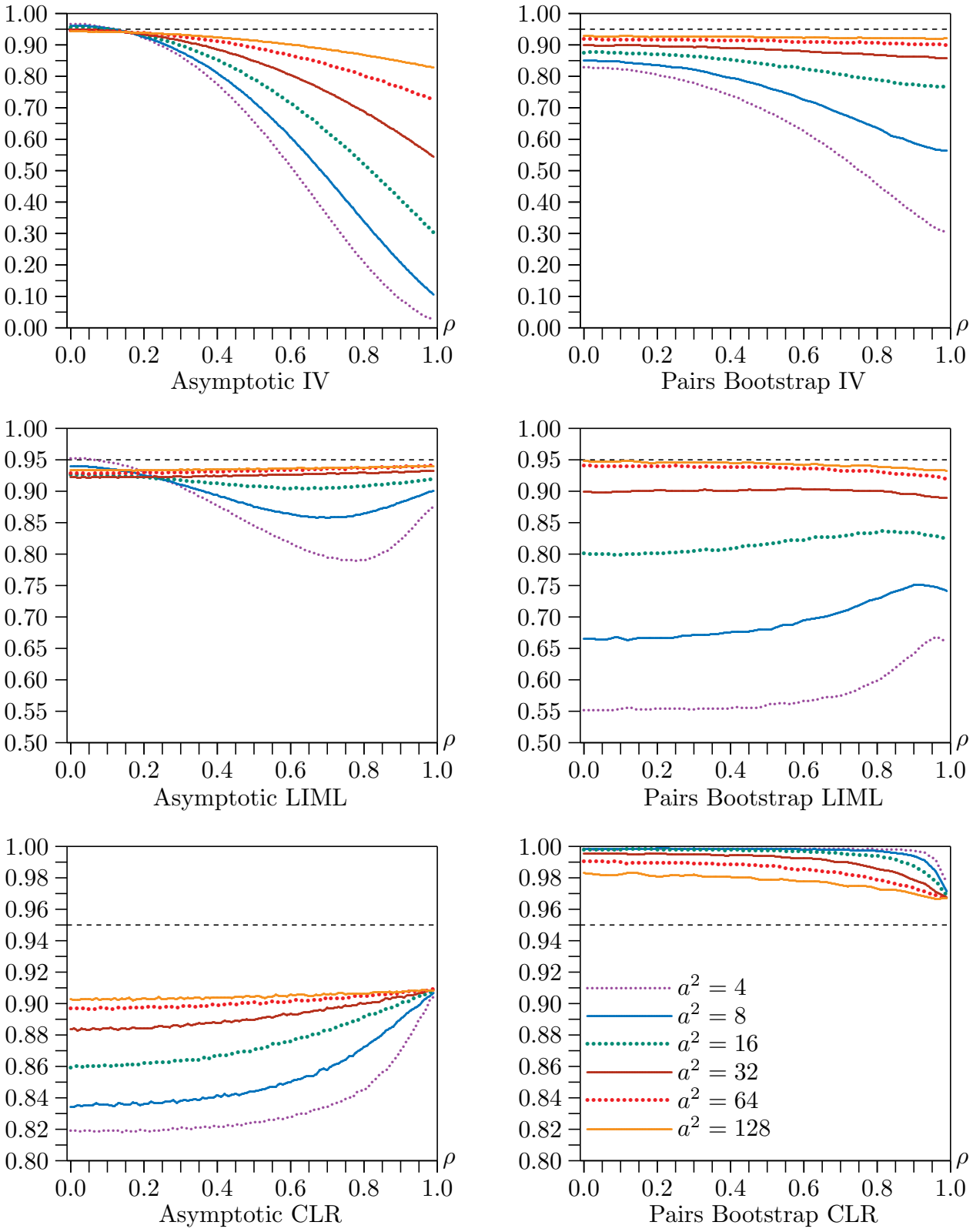

Figure 3. Coverage of confidence sets as functions of ρ for $l - k = 10$ and $n = 50$

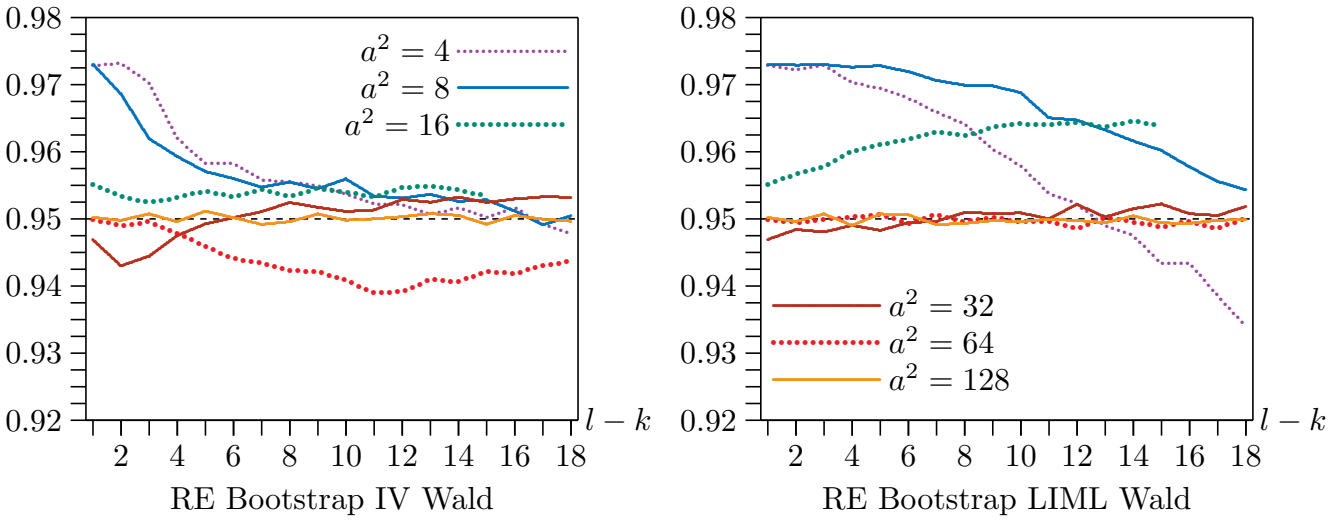

Figure 4. Coverage of confidence sets for $\rho = 0.80$ and $n = 400$

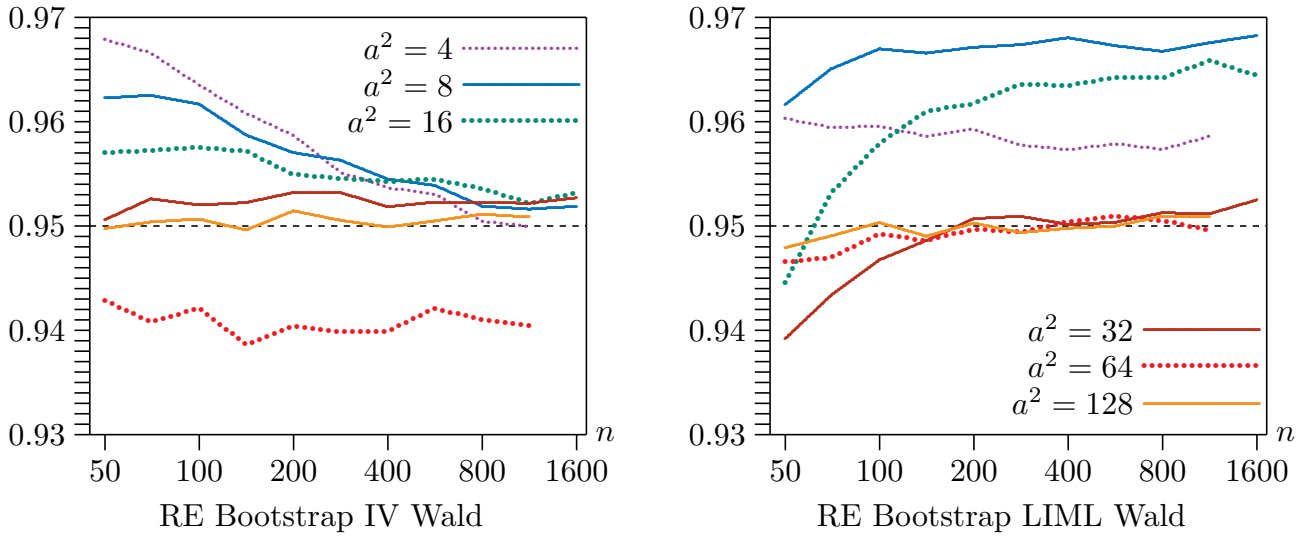

Figure 5. Coverage of confidence sets for $\rho = 0.80$ and $l - k = 10$

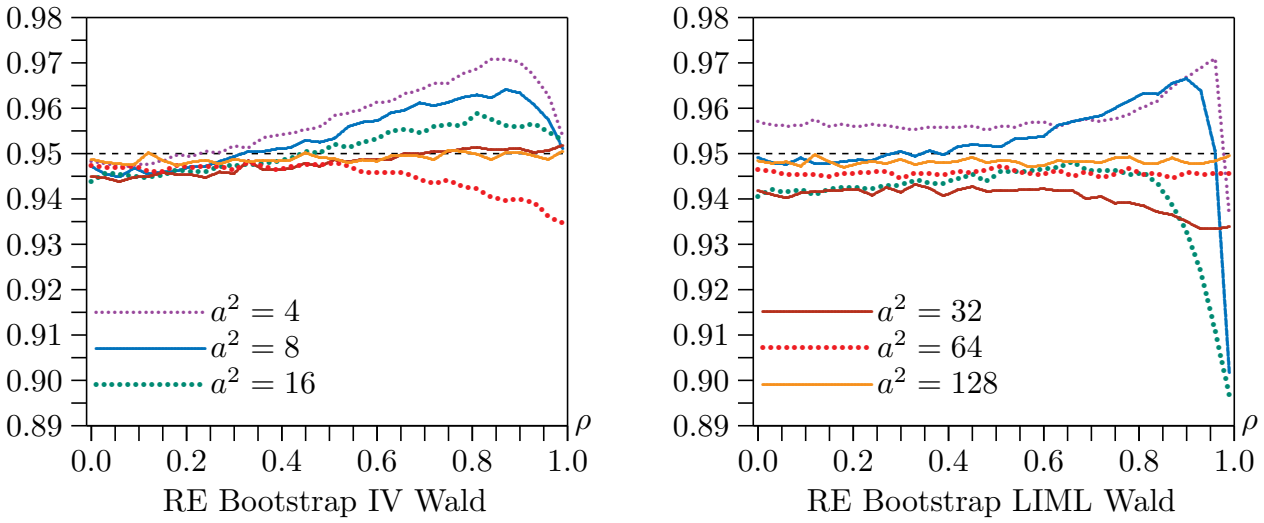

Figure 6. Coverage of confidence sets as functions of ρ for $l - k = 10$ and $n = 50$

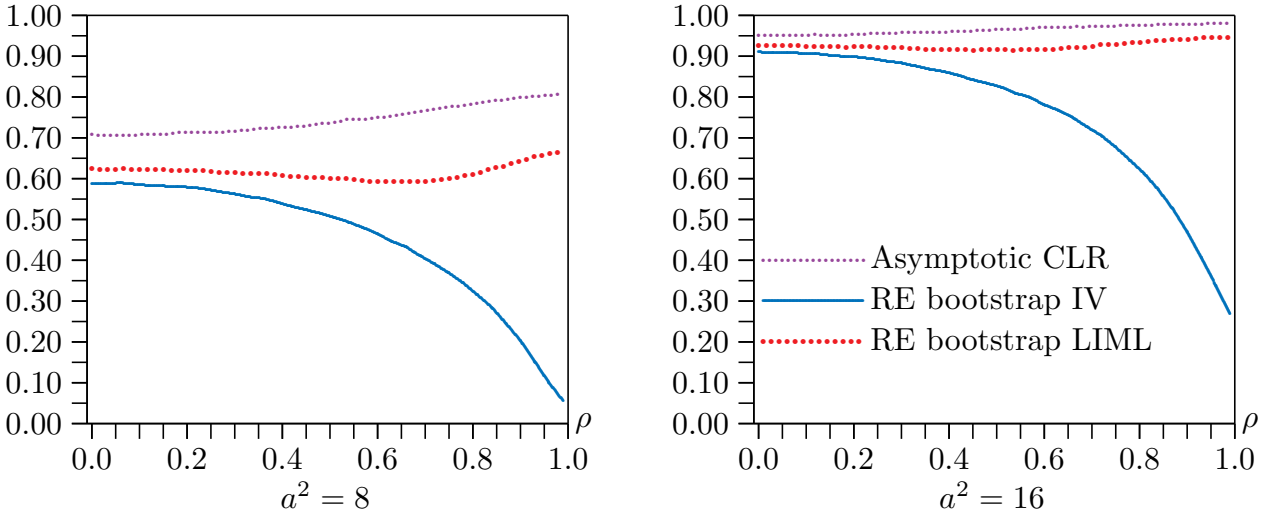

Figure 7. Fraction of bounded intervals as a function of ρ for $l - k = 2$ and $n = 400$

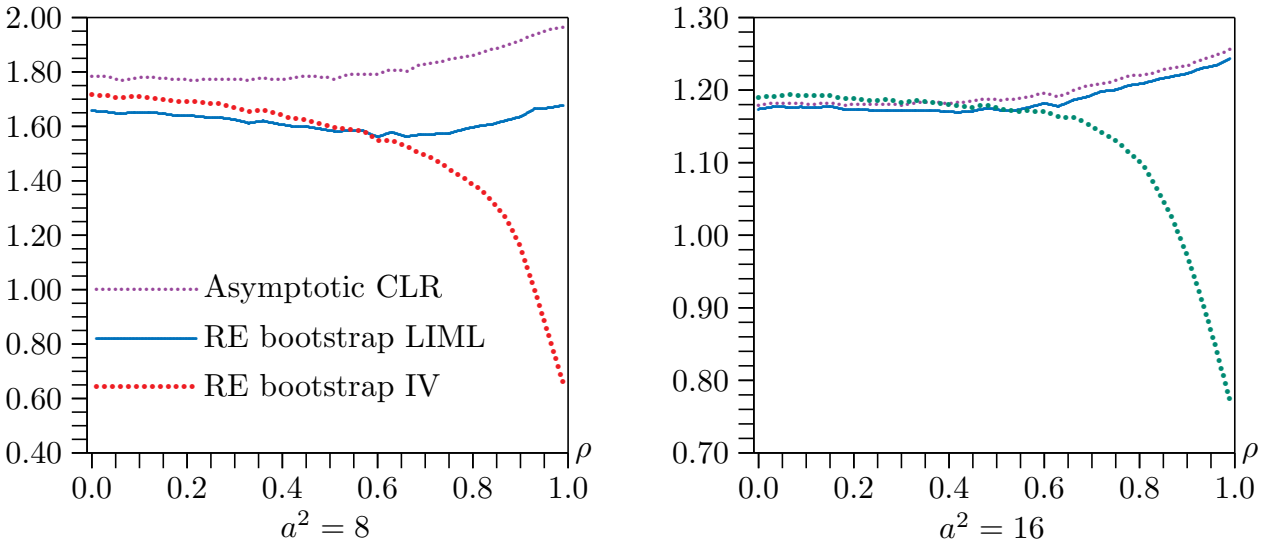

Figure 8. Median length of bounded intervals as functions of ρ for $l - k = 2$ and $n = 400$