

Confidence Sets Based on Inverting Anderson-Rubin Tests

by

Russell Davidson

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

AMSE-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

Economists are often interested in the coefficient of a single endogenous explanatory variable in a linear simultaneous-equations model. One way to obtain a confidence set for this coefficient is to invert the Anderson-Rubin test. The “AR confidence sets” that result have correct coverage under classical assumptions. However, AR confidence sets also have many undesirable properties. It is well known that they can be unbounded when the instruments are weak, as is true of any test with correct coverage. But, even when they are bounded, their length may be very misleading, and their coverage conditional on quantities that the investigator can observe, notably the Sargan statistic for overidentifying restrictions, can be far from correct. A similar property manifests itself, for similar reasons, when a confidence set for a single parameter is based on inverting an F test for two or more parameters.

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.

February 2013

1. Introduction

Classical confidence intervals are, at least implicitly, defined by “inverting” a test. A confidence set at level $1 - \alpha$, which may or may not be a single bounded interval, is simply the set of parameter values for which a test at level α does not reject the null hypothesis that each value in the set is the correct one. This seems to imply that inverting an exact test must lead to a confidence set that has good properties. However, as we show in this paper, that is not the case when the test statistic involves more restrictions than the dimension of the confidence set.

Rather than attempting to state and prove a general result, the paper deals with two special cases. The first is inverting an F test of two or more restrictions to obtain a one-dimensional confidence interval. This is not something that any sensible econometrician would do, of course, but it shows just what the issues are in a very simple context.

The main focus of the paper is confidence sets obtained by inverting the test proposed by Anderson and Rubin (1949). In the linear simultaneous-equations model with weak instruments, the asymptotic distributions of t statistics often provide poor guides to their finite-sample distributions; see Staiger and Stock (1997). As a consequence, confidence intervals based on inverting t tests often have very poor coverage. One proposed solution to this problem is to invert a test which has better finite-sample properties. In a model with just one right-hand-side endogenous variable, the Anderson-Rubin test for the value of the coefficient of that variable is exact under classical assumptions. It has therefore been suggested in several papers, including Dufour (1997), Zivot, Startz, and Nelson (1998), and Dufour and Taamouti (2005), that one should invert the AR test to produce what we shall refer to as an “AR confidence set.”

In this paper, we argue that, although AR confidence sets have correct unconditional coverage, at least under classical assumptions, they have many undesirable properties. Although some of these properties have previously been studied, notably by Zivot, Startz, and Nelson (1998) and Mikusheva (2010), we offer some new theoretical results together with supporting simulation evidence. AR confidence sets do not have correct coverage conditional on the type of confidence set that actually occurs. Moreover, when they are bounded, their length depends on the value of the Sargan statistic for the validity of the overidentifying restrictions. Therefore, any AR confidence set that is actually observed does not have correct coverage. It can be empty, misleadingly short, misleadingly long, or unbounded.

Having correct coverage unconditionally, while desirable, is by itself not very useful. One can always create a $(1 - \alpha)\%$ confidence set with the correct unconditional coverage by setting it equal to the empty set with probability α and the real line with probability $1 - \alpha$. But such a straw-man confidence set provides no useful information. Unfortunately, when the instruments are weak, the AR confidence set may not be much more informative than this straw-man one. Even when they are strong, it never has the correct conditional coverage.

Forchini and Hillier (2003) have argued that the AR statistic is not in fact pivotal, because it does not depend on the parameter of interest everywhere in the parameter space, and that confidence sets based on it are therefore invalid. Our paper is concerned with the more detailed properties of AR confidence sets, but some of the issues that arise below are related to this important point.

It is well known that AR confidence sets may be unbounded. In general, 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). Thus the possible unboundedness of AR confidence sets can actually be seen as a positive feature. What is less widely appreciated is that AR confidence sets may be empty or extremely small. They can thus provide a very misleading impression of how much information the sample provides about the parameter of interest.

The problem of confidence sets that are empty or very small can arise whenever we invert a test that has more degrees of freedom than the number of parameters in which we are interested. In the next section, we show that it can occur when we invert an F test in the classical normal linear regression model. In Section 3, we introduce Anderson-Rubin confidence sets and show that there are four types of them. In Section 4, we explore the important relationship between AR confidence sets and the Sargan statistic for overidentification. In Section 5, we use simulation experiments to study the properties of AR confidence sets. In Section 6, we briefly discuss alternative ways of forming confidence sets in regression models estimated by instrumental variables. Section 7 concludes.

2. Inverting the F Test

The fundamental problem with AR confidence sets is that they are obtained by inverting a test statistic with more than one degree of freedom. A simpler example of the same problem arises in the context of the classical normal linear model

$$\mathbf{y} = \mathbf{x}\beta + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{u}, \quad \mathbf{u} \sim \text{N}(\mathbf{0}, \sigma^2\mathbf{I}), \quad (1)$$

where \mathbf{y} and \mathbf{x} are $n \times 1$ vectors, \mathbf{X}_2 is an $n \times k_2$ matrix, and \mathbf{Z} is an $n \times k_3$ matrix. We assume that the $k \equiv k_2 + k_3 + 1$ columns of \mathbf{x} , \mathbf{X}_2 , and \mathbf{Z} are linearly independent. Suppose we attempt to construct a confidence set for β by inverting the F test for the joint hypothesis

$$\text{H}(\beta_0) : \quad \beta = \beta_0; \quad \boldsymbol{\beta}_2 = \mathbf{0},$$

assuming of course that the true $\boldsymbol{\beta}_2$ is indeed zero. The null model can be written as

$$\mathbf{y} - \mathbf{x}\beta_0 = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{u},$$

and the alternative as

$$\mathbf{y} - \mathbf{x}\beta_0 = \mathbf{X}\boldsymbol{\delta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{u}, \quad (2)$$

where $\mathbf{X} \equiv [\mathbf{x} \ \mathbf{X}_2]$. Clearly, (1) and (2) are just different parametrizations of the same model. The F statistic for a test of $H(\beta_0)$ at nominal level α is

$$F(\beta_0) = \frac{\|\mathbf{P}_{M_Z \mathbf{X}}(\mathbf{y} - \mathbf{x}\beta_0)\|^2/(k_2 + 1)}{\|\mathbf{M}_{[\mathbf{X} \ \mathbf{Z}]}\mathbf{y}\|^2/(n - k)}. \quad (3)$$

Any value of β_0 for which $F(\beta_0) \leq q$, where q is the $1 - \alpha$ quantile of the $F_{k_2+1, n-k}$ distribution, belongs to the confidence set formed by inverting the F test.

Here and throughout the paper, \mathbf{P}_B is for any matrix B the orthogonal projection $B(B^\top B)^{-1}B^\top$ on to the columns of B , and $\mathbf{M}_B \equiv \mathbf{I} - \mathbf{P}_B$ is the complementary orthogonal projection on to the orthogonal complement of the space spanned by the columns of B ; for details see Davidson and MacKinnon (2004), section 2.3. These matrices are symmetric and idempotent.

The inequality $F(\beta_0) \leq q$ can be expressed as a quadratic inequality in β_0 :

$$(\mathbf{x}^\top \mathbf{P}_{M_Z \mathbf{X}} \mathbf{x})\beta_0^2 - 2(\mathbf{x}^\top \mathbf{P}_{M_Z \mathbf{X}} \mathbf{y})\beta_0 + \mathbf{y}^\top (\mathbf{P}_{M_Z \mathbf{X}} - c\mathbf{M}_{[\mathbf{X} \ \mathbf{Z}]})\mathbf{y} \leq 0, \quad (4)$$

where $c \equiv (k_2 + 1)q/(n - k)$. The discriminant of the quadratic is

$$\Delta \equiv 4((\mathbf{x}^\top \mathbf{P}_{M_Z \mathbf{X}} \mathbf{y})^2 - \mathbf{x}^\top \mathbf{P}_{M_Z \mathbf{X}} \mathbf{x} \mathbf{y}^\top (\mathbf{P}_{M_Z \mathbf{X}} - c\mathbf{M}_{[\mathbf{X} \ \mathbf{Z}]})\mathbf{y}). \quad (5)$$

The probability that $\Delta < 0$ is the probability of obtaining an empty confidence set, because the coefficient of β_0^2 in (4) is always positive. Therefore, if the corresponding quadratic equation has no real roots, the quadratic function is everywhere positive, and the inequality is satisfied nowhere. This probability is, of course, less than α .

From (5), we see that $\Delta < 0$ if and only if

$$\mathbf{y}^\top (\mathbf{P}_{M_Z \mathbf{X}} - c\mathbf{M}_{[\mathbf{X} \ \mathbf{Z}]})\mathbf{y} > \frac{(\mathbf{x}^\top \mathbf{P}_{M_Z \mathbf{X}} \mathbf{y})^2}{\mathbf{x}^\top \mathbf{P}_{M_Z \mathbf{X}} \mathbf{x}}. \quad (6)$$

The right-hand side of this inequality is the squared norm of the projection of \mathbf{y} on to the direction of $\mathbf{P}_{M_Z \mathbf{X}} \mathbf{x}$. But, since $\mathbf{P}_{M_Z \mathbf{X}} \mathbf{x} = \mathbf{P}_{M_Z \mathbf{X}} \mathbf{M}_Z \mathbf{x} = \mathbf{M}_Z \mathbf{x}$, the right-hand side of (6) is simply $\mathbf{y}^\top \mathbf{P}_{M_Z \mathbf{x}} \mathbf{y}$.

If we subtract $\mathbf{y}^\top \mathbf{P}_{M_Z \mathbf{x}} \mathbf{y}$ from both sides of (6), the first term inside the parentheses on the left-hand side becomes $\mathbf{P}_{M_Z \mathbf{X}} - \mathbf{P}_{M_Z \mathbf{x}}$. Since

$$\mathbf{P}_{M_Z \mathbf{X}} = \mathbf{P}_{M_Z \mathbf{x}} + \mathbf{P}_{M_{[\mathbf{x} \ \mathbf{Z}]}\mathbf{X}_2}, \quad (7)$$

the inequality (6) can then be rewritten as

$$\mathbf{y}^\top (\mathbf{P}_{M_{[\mathbf{x} \ \mathbf{Z}]}\mathbf{X}_2} - c\mathbf{M}_{[\mathbf{X} \ \mathbf{Z}]})\mathbf{y} > 0,$$

which can be rearranged as

$$\frac{\mathbf{y}^\top \mathbf{P}_{M_{[\mathbf{x} \ \mathbf{Z}]}\mathbf{X}_2} \mathbf{y}/k_2}{\mathbf{y}^\top \mathbf{M}_{[\mathbf{X} \ \mathbf{Z}]}\mathbf{y}/(n - k)} > \left(1 + \frac{1}{k_2}\right)q. \quad (8)$$

The left-hand side of this inequality is distributed as $F_{k_2, n-k}$, and so the probability that $\Delta < 0$ can readily be calculated. The numerical value depends on the nominal coverage $1 - \alpha$, the sample size n , and the numbers k and k_2 of regressors in the model (1).

Suppose without loss of generality that the true value of β is zero and the true value of σ is one. Then the confidence set covers zero if and only if it is non-empty, that is, $\Delta > 0$, and the two real roots of the quadratic have opposite signs. The product of the roots is the ratio of the last term on the left-hand side of (4) to the coefficient of β_0^2 . Since the latter is always positive, the roots have opposite signs if and only if

$$\mathbf{y}^\top (\mathbf{P}_{M_Z \mathbf{x}} - c \mathbf{M}_{[\mathbf{X} \ \mathbf{z}]}) \mathbf{y} < 0, \quad (9)$$

since this inequality implies that $\Delta > 0$; compare (6). The inequality (9) can be rewritten as

$$\frac{\mathbf{y}^\top \mathbf{P}_{M_Z \mathbf{x}} \mathbf{y} / (k_2 + 1)}{\mathbf{y}^\top \mathbf{M}_{[\mathbf{X} \ \mathbf{z}]} \mathbf{y} / (n - k)} < q, \quad (10)$$

and the probability that the inequality is satisfied is of course just $1 - \alpha$, since the left-hand side of (10) is distributed as $F_{k_2+1, n-k}$.

Consider next the statistic for the F test of the part of $H(\beta_0)$ that has nothing to do with β_0 , namely, that $\beta_2 = \mathbf{0}$. This statistic is

$$F_2 \equiv \frac{\mathbf{y}^\top \mathbf{P}_{M_{[\mathbf{x} \ \mathbf{z}]} \mathbf{x}_2} \mathbf{y} / k_2}{\mathbf{y}^\top \mathbf{M}_{[\mathbf{X} \ \mathbf{z}]} \mathbf{y} / (n - k)}. \quad (11)$$

From (7), the left-hand side of (10) can be rewritten as

$$\frac{k_2}{k_2 + 1} F_2 + \frac{\mathbf{y}^\top \mathbf{P}_{M_Z \mathbf{x}} \mathbf{y} / (k_2 + 1)}{\mathbf{y}^\top \mathbf{M}_{[\mathbf{X} \ \mathbf{z}]} \mathbf{y} / (n - k)},$$

and so, if we write $s^2 = \mathbf{y}^\top \mathbf{M}_{[\mathbf{X} \ \mathbf{z}]} \mathbf{y} / (n - k)$, the coverage event can be expressed as

$$k_2 F_2 + \frac{\mathbf{y}^\top \mathbf{P}_{M_Z \mathbf{x}} \mathbf{y}}{s^2} < (k_2 + 1)q,$$

or, equivalently,

$$\mathbf{y}^\top \mathbf{P}_{M_Z \mathbf{x}} \mathbf{y} < s^2 ((k_2 + 1)q - k_2 F_2). \quad (12)$$

The two sides of this inequality are independent, and the left-hand side is distributed as $\chi^2(1)$. Therefore, conditional on F_2 and s^2 , coverage is given by the CDF of $\chi^2(1)$ evaluated at the right-hand side of (12). It is almost never equal to $1 - \alpha$, and the larger is the value of F_2 , the shorter is the interval.

The inequality (12) can never be satisfied if $\Delta < 0$. From (8) and (11), we see that the event $\Delta < 0$ can be written as $k_2 F_2 > (k_2 + 1)q$. It follows that, whenever the

statistic F_2 for $\beta_2 = \mathbf{0}$ is sufficiently large, the confidence interval defined by (4) must be the empty set.

When the confidence interval does exist, its length is the distance between the two roots of the quadratic in (4), that is, $2\sqrt{\Delta}/\mathbf{x}^\top \mathbf{P}_{\mathbf{M}_Z} \mathbf{x}$. It can be seen from (5) and (11) that

$$\Delta = 4\mathbf{x}^\top \mathbf{P}_{\mathbf{M}_Z} \mathbf{x} s^2((k_2 + 1)q - k_2 F_2),$$

and so the length of the interval, when it exists, is

$$\frac{2s((k_2 + 1)q - k_2 F_2)^{1/2}}{(\mathbf{x}^\top \mathbf{P}_{\mathbf{M}_Z} \mathbf{x})^{1/2}}. \quad (13)$$

As noted earlier, the coverage of the confidence interval defined by (4) is given by the CDF of $\chi^2(1)$ evaluated at the right-hand side of (12). From (13), this coverage can also be expressed as the CDF of $\chi^2(1)$ evaluated at 4 times the squared length of the interval multiplied by $\mathbf{x}^\top \mathbf{P}_{\mathbf{M}_Z} \mathbf{x}$.

It is evident from expression (13) that, if $\hat{\beta}_2$ differed substantially from a zero vector, and F_2 were consequently a large number, $(k_2 + 1)q - k_2 F_2$ would be negative, and there would not exist a bounded interval. That could happen either by chance or because $\beta_2 \neq \mathbf{0}$. It seems very unsatisfactory that the length, and even the existence, of a confidence interval for β should depend on the value of β_2 .

Of course, in the context of the classical linear model (1), it makes no sense to base a confidence interval for β on the statistic $F(\beta_0)$ defined in (3). The usual interval is instead based on the t statistic for $\beta = \beta_0$. But, as we shall see in the next section, inverting an AR test is very much like inverting $F(\beta_0)$.

3. Anderson-Rubin Confidence Sets

We deal with the simultaneous two-equation linear model

$$\mathbf{y}_1 = \beta \mathbf{y}_2 + \mathbf{Z}\gamma + \mathbf{u}_1 \quad (14)$$

$$\mathbf{y}_2 = \mathbf{W}\pi + \mathbf{u}_2 = \mathbf{Z}\pi_1 + \mathbf{W}_2\pi_2 + \mathbf{u}_2. \quad (15)$$

Here \mathbf{y}_1 and \mathbf{y}_2 are n -vectors of observations on endogenous variables, \mathbf{Z} is an $n \times k$ matrix of observations on exogenous variables, and \mathbf{W} is an $n \times l$ matrix of exogenous instruments with the property that $\mathcal{S}(\mathbf{Z})$, the subspace spanned by the columns of \mathbf{Z} , lies in $\mathcal{S}(\mathbf{W})$, the subspace spanned by the columns of \mathbf{W} . The $n \times (l - k)$ matrix \mathbf{W}_2 is constructed in such a way that $\mathcal{S}(\mathbf{Z}, \mathbf{W}_2) = \mathcal{S}(\mathbf{W})$, and \mathbf{W} is assumed to have full column rank. Equation (14) is a structural equation, and equation (15) is a reduced-form equation. The model is overidentified if $l > k + 1$, the number of overidentifying restrictions being $l - k - 1$.

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

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

For the AR test to be exact, we also need the disturbances to be normally distributed. The Anderson-Rubin statistic for a test of the hypothesis that $\beta = \beta_0$ is

$$\text{AR}(\beta_0) = \frac{n-l}{l-k} \frac{(\mathbf{y}_1 - \beta_0 \mathbf{y}_2)^\top \mathbf{P}_1 (\mathbf{y}_1 - \beta_0 \mathbf{y}_2)}{(\mathbf{y}_1 - \beta_0 \mathbf{y}_2)^\top \mathbf{M}_W (\mathbf{y}_1 - \beta_0 \mathbf{y}_2)}, \quad (16)$$

where $\mathbf{P}_1 \equiv \mathbf{M}_Z - \mathbf{M}_W = \mathbf{P}_W - \mathbf{P}_Z$. Under the null hypothesis, the AR statistic (16) is distributed as $F(l-k, n-l)$. This statistic is, of course, minimized at the LIML estimator $\hat{\beta}_{\text{LIML}}$.

Let q now denote the $1 - \alpha$ quantile of the $F(l-k, n-l)$ distribution. Then β_0 belongs to the confidence set at level $1 - \alpha$ if and only if $\text{AR}(\beta_0) \leq q$. This inequality can be reformulated as

$$(\mathbf{y}_2^\top \mathbf{A} \mathbf{y}_2) \beta_0^2 - 2(\mathbf{y}_1^\top \mathbf{A} \mathbf{y}_2) \beta_0 + \mathbf{y}_1^\top \mathbf{A} \mathbf{y}_1 \geq 0, \quad (17)$$

where $\mathbf{A} = c \mathbf{M}_W - \mathbf{P}_1$, with $c = q(l-k)/(n-l)$; compare (4). Zivot, Startz, and Nelson (1998) study this inequality in some detail. For convenience, we summarize their results here.

The discriminant of the quadratic equation obtained by replacing the inequality in (17) by an equality is

$$D \equiv 4(\mathbf{y}_1^\top \mathbf{A} \mathbf{y}_2)^2 - 4\mathbf{y}_1^\top \mathbf{A} \mathbf{y}_1 \mathbf{y}_2^\top \mathbf{A} \mathbf{y}_2; \quad (18)$$

compare (5). If $D < 0$, the equation has no real roots, so that the inequality (17) is either always or never satisfied. It is always satisfied if $\mathbf{y}_2^\top \mathbf{A} \mathbf{y}_2 > 0$, and so, in this case, the confidence set is the entire real line. It is never satisfied if $\mathbf{y}_2^\top \mathbf{A} \mathbf{y}_2 < 0$, which implies that the confidence set is empty.

If $D > 0$, the equation has two real roots. If $\mathbf{y}_2^\top \mathbf{A} \mathbf{y}_2 < 0$, the confidence set is the interval between them, while, if $\mathbf{y}_2^\top \mathbf{A} \mathbf{y}_2 > 0$, it is the set composed of the disjoint union of the open infinite interval from the upper root to $+\infty$ and that from the lower root to $-\infty$.

Regardless of the sign of D , in the knife-edge case for which $\mathbf{y}_2^\top \mathbf{A} \mathbf{y}_2 = 0$, the inequality (17) is satisfied with equality when $\beta_0 = \mathbf{y}_1^\top \mathbf{A} \mathbf{y}_1 / (2\mathbf{y}_1^\top \mathbf{A} \mathbf{y}_2)$, and the confidence set extends unboundedly from this value to the right or left according as $\mathbf{y}_1^\top \mathbf{A} \mathbf{y}_2$ is negative or positive. In the doubly knife-edge case with $\mathbf{y}_2^\top \mathbf{A} \mathbf{y}_2 = \mathbf{y}_1^\top \mathbf{A} \mathbf{y}_2 = 0$, the confidence set is the whole real line if $\mathbf{y}_1^\top \mathbf{A} \mathbf{y}_1 \geq 0$, and is empty if $\mathbf{y}_1^\top \mathbf{A} \mathbf{y}_1 < 0$.

It follows that the confidence set is unbounded whenever $\mathbf{y}_2^\top \mathbf{A} \mathbf{y}_2 \geq 0$. This condition can be rewritten as

$$c \mathbf{y}_2^\top \mathbf{M}_W \mathbf{y}_2 - \mathbf{y}_2^\top \mathbf{P}_1 \mathbf{y}_2 \geq 0.$$

Using the definition of c and a little algebra allows us to rewrite this inequality as

$$\frac{\mathbf{y}_2^\top \mathbf{P}_1 \mathbf{y}_2 / (l-k)}{\mathbf{y}_2^\top \mathbf{M}_W \mathbf{y}_2 / (n-l)} \leq q. \quad (19)$$

The quantity on the left-hand side of (19) is the ordinary F statistic for $\boldsymbol{\pi}_2 = \mathbf{0}$ in equation (15), and q is the critical value for a test at level α based on this statistic. Thus, as Zivot, Startz, and Nelson (1998) showed, the the AR confidence set is unbounded whenever the statistic for $\boldsymbol{\pi}_2 = \mathbf{0}$ in (15) is less than q , that is, whenever we cannot reject the hypothesis that the instruments that are not also explanatory variables (namely, the columns of \mathbf{W}_2) have no explanatory power for \mathbf{y}_2 .

It should be noted that the confidence set cannot be empty if the model is just identified. In that case, the LIML estimator $\hat{\beta}_{\text{LIML}}$ is equal to the instrumental-variables (IV) estimator $\hat{\beta}_{\text{IV}}$, which satisfies the estimating equation $\mathbf{y}_2^\top \mathbf{P}_1 (\mathbf{y}_1 - \mathbf{y}_2 \hat{\beta}_{\text{IV}}) = 0$. In this case, the image of \mathbf{P}_1 is one-dimensional. Thus the vector $\mathbf{y}_1 - \mathbf{y}_2 \hat{\beta}_{\text{LIML}}$ is orthogonal to all vectors in that image, and we see from (16) that $\text{AR}(\hat{\beta}_{\text{LIML}})$ is zero. Consequently, $\hat{\beta}_{\text{LIML}}$ always belongs to the confidence set in the just-identified case, whatever the significance level.

There is no point calculating an AR confidence set whenever the inequality (19) holds, because a set that consists of the entire real line, perhaps with a hole in the middle, tells us nothing useful about the value of β . In contrast to the confidence set, the identifiability test statistic does provide valuable information, since it provides a natural measure of the strength of the instruments; see Stock and Yogo (2005).

The fact that some types of AR and other confidence set are unbounded when the instruments are sufficiently weak can be viewed as a consequence of a fundamental result of Dufour (1997), who showed that no valid confidence set which is almost surely bounded exists in the neighborhood of a point where the parameter is not identified.

Unconditionally, the AR confidence set always has the correct coverage. However, once we observe what type of set it is, that is no longer the case. By construction, the empty set undercovers, and the real line overcovers. The bounded interval and the disjoint interval can either overcover or undercover. This is of course the case for any confidence set that may be empty or the whole real line.

4. Relations with the Sargan Test

The Sargan statistic for overidentifying restrictions (Sargan, 1958) should be computed whenever $l - k - 1 > 0$, that is, unless the system is just identified. It is most commonly computed as $1/\hat{\sigma}_1^2$ times the minimized value of the IV criterion function,

$$\frac{1}{\hat{\sigma}_1^2} (\mathbf{y}_1 - \hat{\beta}_{\text{IV}} \mathbf{y}_2)^\top \mathbf{P}_W (\mathbf{y}_1 - \hat{\beta}_{\text{IV}} \mathbf{y}_2) = \frac{1}{\hat{\sigma}_1^2} (\mathbf{y}_1 - \hat{\beta}_{\text{IV}} \mathbf{y}_2)^\top \mathbf{P}_1 (\mathbf{y}_1 - \hat{\beta}_{\text{IV}} \mathbf{y}_2), \quad (20)$$

where $\hat{\beta}_{\text{IV}}$ is the IV (or two-stage least squares) estimate of β , and the estimated variance $\hat{\sigma}_1^2$ denotes $n^{-1} \hat{\mathbf{u}}_1^\top \mathbf{M}_Z \hat{\mathbf{u}}_1$, with $\hat{\mathbf{u}}_1 \equiv \mathbf{y}_1 - \hat{\beta}_{\text{IV}} \mathbf{y}_2$. The equality in (20) follows from the fact that

$$(\mathbf{M}_Z - \mathbf{M}_W)(\mathbf{y}_1 - \hat{\beta}_{\text{IV}} \mathbf{y}_2) = (\mathbf{I} - \mathbf{M}_W)(\mathbf{y}_1 - \hat{\beta}_{\text{IV}} \mathbf{y}_2) = \mathbf{P}_W (\mathbf{y}_1 - \hat{\beta}_{\text{IV}} \mathbf{y}_2),$$

because \mathbf{Z} is orthogonal to the IV residuals.

It is evident that the numerator of the expression on the right-hand side of equation (20) would be identical to the numerator of the AR statistic (16) if $\hat{\beta}_{\text{IV}}$ were replaced by β_0 . The latter will always be no smaller than the former, because $\hat{\beta}_{\text{IV}}$ minimizes the numerator. That is why the AR statistic has $l - k$ degrees of freedom in the numerator, while the Sargan statistic (which, of course, is not exact) has $l - k - 1$. It is not hard to show that the numerator of (16) can be rewritten as

$$(\mathbf{y}_1 - \hat{\beta}_{\text{IV}}\mathbf{y}_2)^\top \mathbf{P}_1(\mathbf{y}_1 - \hat{\beta}_{\text{IV}}\mathbf{y}_2) + (\hat{\beta}_{\text{IV}}\mathbf{y}_2 - \beta_0\mathbf{y}_2)^\top \mathbf{P}_1(\hat{\beta}_{\text{IV}}\mathbf{y}_2 - \beta_0\mathbf{y}_2). \quad (21)$$

The first term in (21) is the numerator of the Sargan statistic (20). Thus, if the Sargan and AR statistics had the same denominator, the latter would always be no smaller than the former. This is not always true in finite samples, because the denominators are not the same, although they both estimate σ_1^2 consistently under the null. But there is inevitably a very strong tendency for large values of the Sargan statistic to be associated with large values of the AR statistic.

In order to analyze the statistical properties of the AR confidence set and its relationship to the Sargan statistic, we need to specify a data-generating process. Following Davidson and MacKinnon (2008), we use the DGP:

$$\begin{aligned} \mathbf{y}_1 &= \beta\mathbf{y}_2 + \mathbf{u}_1, \\ \mathbf{y}_2 &= a\mathbf{w} + \mathbf{u}_2, \end{aligned} \quad (22)$$

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

$$\begin{aligned} \mathbf{u}_1 &= r\mathbf{v}_1 + \rho\mathbf{v}_2, \\ \mathbf{u}_2 &= \mathbf{v}_2, \end{aligned} \quad \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix} \sim \text{N}(\mathbf{0}, \mathbf{I}), \quad r^2 + \rho^2 = 1. \quad (23)$$

This DGP is a special case of the more general model specified by (14) and (15). However, by varying the three parameters β , ρ , and a , we can generate AR or Sargan statistics with the same distributions as those generated by any DGP contained in the more general model; see section 3 of Davidson and MacKinnon (2008).

The vector w_1 lies in the direction of $\mathbf{W}\boldsymbol{\pi}$ in (15). It provides all the explanatory power of all the instruments in the matrix \mathbf{W} . Since it is only $\mathcal{S}(\mathbf{W})$ that matters, we are perfectly free to perform a linear transformation on \mathbf{W} that makes this the case. Similar reasoning shows that we may ignore any exogenous covariates in the matrix \mathbf{Z} by the expedient of replacing all other variables by the respective residuals from regressing them linearly on \mathbf{Z} , that is, by pre-multiplying each variable by $\mathbf{M}_{\mathbf{Z}}$.

By normalizing the vector \mathbf{w} to have squared length unity, 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 this parameter is called the scalar concentration parameter; see Phillips (1983, p. 470) and Stock, Wright, and Yogo (2002).

Finally, the fact that we can without loss of generality set the variances σ_1^2 and σ_2^2 in (22) to unity is a consequence of the fact that all the statistics we consider are homogeneous of degree zero with respect to \mathbf{y}_1 and \mathbf{y}_2 separately.

As we have seen, the AR confidence set is a bounded interval if and only if $D > 0$ and $\mathbf{y}_2^\top \mathbf{A} \mathbf{y}_2 < 0$. In this case, the length of the interval is the distance between the two roots of the quadratic in (17), which is $-\sqrt{D}/\mathbf{y}_2^\top \mathbf{A} \mathbf{y}_2$. Under the DGP given by (22) and (23), the limit of this ratio as $a \rightarrow \infty$ is zero. The quantity that has a non-trivial limit as $a \rightarrow \infty$ is thus the length of the interval times a . It can be shown that this limit is the square root of

$$c(\mathbf{y}_1 - \beta \mathbf{y}_2)^\top \mathbf{M}_{\mathbf{W}}(\mathbf{y}_1 - \beta \mathbf{y}_2) - (\mathbf{y}_1 - \beta \mathbf{y}_2)^\top (\mathbf{P}_1 - \mathbf{P}_{\mathbf{w}})(\mathbf{y}_1 - \beta \mathbf{y}_2). \quad (24)$$

The first term in (24) is c times a random variable that follows the $\chi^2(n-l)$ distribution. The second term is an independent random variable that follows the $\chi^2(l-k-1)$ distribution. Of course, both these quantities would have to be divided by σ_1^2 if we had not set it to unity. The distribution of the second term, and its independence from the first term, both follow from the fact that the matrix $\mathbf{P}_1 - \mathbf{P}_{\mathbf{w}} = \mathbf{P}_{\mathbf{W}} - \mathbf{P}_{\mathbf{Z}} - \mathbf{P}_{\mathbf{w}}$ projects onto the $l-k-1$ components of \mathbf{W} that do not lie in $\mathcal{S}(\mathbf{w}, \mathbf{Z})$.

Expression (24) is random and may be either positive or negative. It is most likely to be negative when α is large, so that q , the $1-\alpha$ quantile of $F(l-k, n-l)$, is small. There is evidently a non-empty confidence set only when it is positive. Since we are considering the limit as $a \rightarrow \infty$, there is no danger that the set will be unbounded.

It is interesting to see how expression (24) is related to a slightly modified version of the Sargan statistic (20). The modified statistic is

$$S = \frac{\hat{\mathbf{u}}_1^\top \mathbf{P}_1 \hat{\mathbf{u}}_1}{\check{\sigma}_1^2}, \quad \check{\sigma}_1^2 = \frac{1}{n-l} \hat{\mathbf{u}}_1^\top \mathbf{M}_{\mathbf{W}} \hat{\mathbf{u}}_1, \quad (25)$$

where again $\hat{\mathbf{u}}_1 = \mathbf{y}_1 - \hat{\beta}_{\text{IV}} \mathbf{y}_2$. This differs from (20) because, instead of using the usual variance estimate $\hat{\sigma}_1^2$, it uses the same one as the AR statistic for testing $\beta = \hat{\beta}_{\text{IV}}$.

The numerator of S is

$$\hat{\mathbf{u}}_1^\top \mathbf{P}_1 \hat{\mathbf{u}}_1 = (\mathbf{y}_1 - \hat{\beta}_{\text{IV}} \mathbf{y}_2)^\top \mathbf{P}_1 (\mathbf{y}_1 - \hat{\beta}_{\text{IV}} \mathbf{y}_2) = \mathbf{y}_1^\top \mathbf{P}_1 (\mathbf{y}_1 - \hat{\beta}_{\text{IV}} \mathbf{y}_2), \quad (26)$$

where the second equality follows from the estimating equation $\mathbf{y}_2^\top \mathbf{P}_1 (\mathbf{y}_1 - \hat{\beta}_{\text{IV}} \mathbf{y}_2) = 0$ that defines $\hat{\beta}_{\text{IV}}$. This equation implies that

$$\hat{\beta}_{\text{IV}} = \frac{\mathbf{y}_2^\top \mathbf{P}_1 \mathbf{y}_1}{\mathbf{y}_2^\top \mathbf{P}_1 \mathbf{y}_2}. \quad (27)$$

Substituting (27) into the rightmost expression in (26) yields

$$\begin{aligned} \hat{\mathbf{u}}_1^\top \mathbf{P}_1 \hat{\mathbf{u}}_1 &= \mathbf{y}_1^\top \mathbf{P}_1 \mathbf{y}_1 - \mathbf{y}_1^\top \mathbf{P}_1 \mathbf{y}_2 (\mathbf{y}_2^\top \mathbf{P}_1 \mathbf{y}_2)^{-1} \mathbf{y}_2^\top \mathbf{P}_1 \mathbf{y}_1 \\ &= \mathbf{y}_1^\top (\mathbf{P}_1 - \mathbf{P}_{\mathbf{P}_1 \mathbf{y}_2}) \mathbf{y}_1 = \mathbf{u}_1^\top (\mathbf{P}_1 - \mathbf{P}_{\mathbf{P}_1 \mathbf{y}_2}) \mathbf{u}_1. \end{aligned}$$

Thus from (25) we have

$$\check{\sigma}_1^2 S = \mathbf{u}_1^\top (\mathbf{P}_1 - \mathbf{P}_{\mathbf{P}_1 \mathbf{y}_2}) \mathbf{u}_1 = \mathbf{u}_1^\top \mathbf{P}_1 \mathbf{u}_1 - \frac{(\mathbf{u}_1^\top \mathbf{P}_1 \mathbf{y}_2)^2}{\mathbf{y}_2^\top \mathbf{P}_1 \mathbf{y}_2}. \quad (28)$$

If we replace \mathbf{y}_2 by $a\mathbf{w} + \mathbf{u}_2$ and retain only the leading-order terms as $a \rightarrow \infty$, the term that is subtracted in the rightmost expression here tends to $(\mathbf{w}^\top \mathbf{u}_1)^2 = \mathbf{u}_1^\top \mathbf{P}_w \mathbf{u}_1$, where the equality follows from the fact that $\mathbf{w}^\top \mathbf{w} = 1$. Thus, in the limit,

$$\check{\sigma}_1^2 S = \mathbf{u}_1^\top (\mathbf{P}_1 - \mathbf{P}_w) \mathbf{u}_1 = (\mathbf{y}_1 - \beta \mathbf{y}_2)^\top (\mathbf{P}_1 - \mathbf{P}_w) (\mathbf{y}_1 - \beta \mathbf{y}_2). \quad (29)$$

It is easy to see that $\hat{\beta}_{\text{IV}}$ tends to β as $a \rightarrow \infty$. From (27),

$$\hat{\beta}_{\text{IV}} = \frac{(a\mathbf{w} + \mathbf{u}_2)^\top \mathbf{P}_1 (\beta(a\mathbf{w} + \mathbf{u}_2) + \mathbf{u}_1)}{(a\mathbf{w} + \mathbf{u}_2)^\top \mathbf{P}_1 (a\mathbf{w} + \mathbf{u}_2)} = \beta + \frac{(a\mathbf{w} + \mathbf{u}_2)^\top \mathbf{P}_1 \mathbf{u}_1}{(a\mathbf{w} + \mathbf{u}_2)^\top \mathbf{P}_1 (a\mathbf{w} + \mathbf{u}_2)}.$$

Since the second term in the rightmost expression here is $O(a)/O(a^2) = O(a^{-1})$, that expression vanishes as $a \rightarrow \infty$. The consistency of $\hat{\beta}_{\text{IV}}$ implies that

$$\hat{\mathbf{u}}_1^\top \mathbf{M}_w \hat{\mathbf{u}}_1 / (n - l) = (\mathbf{y}_1 - \beta \mathbf{y}_2)^\top \mathbf{M}_w (\mathbf{y}_1 - \beta \mathbf{y}_2) / (n - l) + O(a^{-1})$$

as $a \rightarrow \infty$. Thus the first term in (24) can be replaced by

$$q(l - k) \frac{\hat{\mathbf{u}}_1^\top \mathbf{M}_w \hat{\mathbf{u}}_1}{n - l} = q(l - k) \check{\sigma}_1^2.$$

Similarly, by (29), the second term can be replaced by $\check{\sigma}_1^2 S$. We conclude that, in the limit as $a \rightarrow \infty$, the length of the bounded AR interval, if it exists, is simply

$$\check{\sigma}_1 (q(l - k) - S)^{1/2}. \quad (30)$$

This is a deterministic function of $\check{\sigma}_1$ and S , which is proportional to the former and nonlinear in the latter. As S increases, the interval becomes shorter and eventually ceases to exist.

Although the result (30) is strictly true only in the limit, it may be expected to provide a good guide whenever a is reasonably large, that is, whenever the instruments are reasonably strong. It implies that, when the AR confidence set is a bounded interval, its coverage will vary inversely with the magnitude of the Sargan statistic. This may be especially problematic in practice if, as will very often be the case, the overidentifying restrictions are not quite satisfied. In consequence, observed Sargan statistics may well tend to be larger than they should be by chance, and bounded AR intervals consequently shorter.

The fundamental reason for the result that the AR confidence set depends on the value of the Sargan statistic is that the AR statistic has more than one degree of

freedom. The Sargan statistic plays exactly the same role in (30) as k_2 times the statistic F_2 did in (13). In obtaining (13), there was no need to consider a limiting argument. The only reason we needed a limiting argument to obtain (30) is that the Sargan statistic does not have a unique distribution independent of the model parameters when a is finite.

Figure 1 illustrates all four types of interval by graphing $\text{AR}(\beta_0)$ against β_0 . The dashed horizontal line is the critical value, q . Two variants of the bounded interval case are shown. In one of these, the interval is very short, and in the other it is quite long. What type of interval we obtain depends on α . In particular, the probability that the interval is an empty set diminishes as α becomes smaller and q consequently becomes larger. All five intervals in the figure are for samples drawn from the same data-generating process, for which the instruments are moderately weak. The figure illustrates the fact that sampling variation can produce radically different AR confidence sets.

The confidence sets considered in Zivot, Startz, and Nelson (1998), which were constructed by inverting the LR test and variants of the LM test, share a number of properties with AR confidence sets. They, too, can be unbounded, and so they have incorrect coverage conditional on being bounded or not. What differentiates AR confidence sets from these others is that, since the AR statistic depends on the Sargan statistic, an AR confidence set can be empty, and, as the figure illustrates, a bounded interval can be very much too short. Thus we cannot interpret an observed AR confidence set, even a bounded interval, in the way we would like to interpret a confidence interval. On average, at least when the model is well identified, bounded intervals must overcover, in order to offset the failure of empty sets to cover at all. But there will always be bounded intervals like the one shown in the top panel of Figure 1 which give the misleading impression that we have estimated β much more accurately than is actually the case.

The AR set in the just-identified case is not subject to this criticism. This is a natural consequence of the fact that, in this case, the Sargan statistic is zero. In fact, it is easy to see that, for a just-identified model, the AR test statistic is equal to the version of the LM statistic proposed by Kleibergen (2002) and Moreira (2009).

5. Properties of AR Confidence Sets

In this section, we use simulation experiments to study various properties of AR confidence sets, including their conditional coverage. We generate artificial data from the DGP specified by (22) and (23). Because this DGP uses weak instrument asymptotics, the sample size does not matter much once it exceeds a threshold size. In Davidson and MacKinnon (2010), we found that the performance of various test statistics for β changed very little once n exceeded 400. We therefore set $n = 400$ in all our experiments. For each DGP, we generated 500,000 simulated datasets.

The key parameters in our experiments are a , ρ , and $l - k$. To save space, we report results only for $l - k = 7$, which means that the model is moderately overidentified.

Results for substantially smaller or larger values of $l - k$ might look quite different, but that would primarily be because a needs to increase with $l - k$ in order to keep the strength of the instruments constant. The basic structure of the results does not seem to change much with $l - k$.

Figure 2 shows how the frequencies of the four types of 95% AR confidence set depend on a and ρ . The figure has four panels, which correspond to four different values of a . The value of ρ , which varies from 0.00 to 0.99 by increments of 0.01, is on the horizontal axis. Negative values are not included, because the figures would simply be symmetric around $\rho = 0$.

When $a = 1$, the instruments are extremely weak, and when $a = 8$ they are minimally strong. In the former case, the 95% AR confidence set is unbounded about 90% of the time. For most values of ρ , the unbounded set is usually the entire real line. However, as ρ becomes larger, the case of two unbounded segments becomes more common, until it almost completely drives out the real-line case when $\rho = 0.99$. The results for $a = 2$ are similar to those for $a = 1$, except that the bounded interval becomes somewhat more common (but it still occurs less than 25% of the time), and the two unbounded sets become somewhat less common.

The results change dramatically when we move from $a = 2$ to $a = 4$. The 95% AR confidence set is now bounded more than 80% of the time, and the empty set is a good deal more common than it was before. Finally, when $a = 8$, there is just a handful of unbounded confidence sets in 50 million replications, and the bounded interval occurs between 97.1% and 97.4% of the time. The empty set occurs very slightly more often as ρ increases.

Figure 3 shows conditional coverage for four types of confidence set for the same experiments as Figure 2. We do not bother to show coverage for the real line or the empty set. Instead, we show it for bounded intervals when the Sargan statistic, computed in the usual way as (20), either exceeds the 0.90 quantile of the $\chi^2(l - k - 1)$ distribution (“ S large”) or falls short of the 0.50 quantile (“ S modest”). Several striking results are apparent from the figure.

- When a is small, the bounded interval may either overcover slightly (when ρ is small) or undercover severely (when ρ is large and $a = 1$). When a is not small, the bounded interval always overcovers, as it must do in order to offset the undercoverage associated with the empty set.
- The two-segment confidence set undercovers when ρ is small. However, as ρ increases, its coverage increases, and it eventually overcovers. This type of confidence set does not occur when $a = 8$.
- The coverage of the bounded interval changes dramatically when we condition on the Sargan statistic. When the latter rejects at the nominal 0.10 level, the bounded interval always undercovers, often severely. In contrast, when it fails to reject at the 0.50 level, the bounded interval always overcovers except for larger values of ρ when $a = 1$. This overcoverage is generally quite extreme.

For example, when $a = 8$, the 95% bounded AR interval always covers at least 99.8% of the time when the Sargan statistic fails to reject at the 0.50 level.

These results suggest that the length of a bounded AR confidence interval will generally provide a poor guide to the precision with which the parameter β is estimated. To investigate this conjecture, we calculated the dispersion of $\hat{\beta}_{\text{LIML}}$ as the difference between its 0.025 and 0.975 quantiles over the 500,000 replications. In Figure 4, we compare this with the median and with the 0.01 and 0.99 quantiles of the lengths of the 95% AR confidence sets when they are bounded intervals. Ideally, the median length of the bounded AR intervals should be very similar to the dispersion of the estimates, and the upper-tail and lower-tail quantiles of interval length should not be too much higher or lower than the median.

The three left-hand panels of Figure 4 show results for 95% AR confidence sets when they are bounded intervals, and the three right-hand panels show results for conventional Wald intervals based on $\hat{\beta}_{\text{LIML}}$. It is appropriate to compare AR intervals with ones based on $\hat{\beta}_{\text{LIML}}$, because the AR statistic is minimized at $\hat{\beta}_{\text{LIML}}$. Results are presented for $a = 4$, $a = 8$, and $a = 16$. We do not present results for smaller values of a because most of the AR confidence sets were unbounded (see Figure 2) and because it is unreasonable to expect any method to produce reliable results in these cases. Note that the vertical axis is logarithmic.

It is evident that the median length of the bounded 95% AR interval is generally a poor guide to the dispersion of $\hat{\beta}_{\text{LIML}}$. The former always overestimates the latter, and the problem does not go away as a becomes larger. Moreover, the length of the bounded AR intervals evidently varies greatly. When $a = 4$, the upper-tail quantile of the distribution of their lengths can be more than 80 times the dispersion of $\hat{\beta}_{\text{LIML}}$, while the lower-tail quantile can be no more than 1/4 of the dispersion. Of course, as the theory of Section 4 makes clear, there are a few bounded intervals that are just barely longer than zero, but these are evidently well to the left of the 0.01 quantile. For large a , this occurs whenever $q(l - k) - S$ in equation (30) is just barely positive.

Whereas the median length of the AR intervals always overstates the dispersion of $\hat{\beta}_{\text{LIML}}$, that of the Wald LIML intervals always understates it (although just by a whisker when $a = 16$). The lengths of the Wald intervals vary much less than those of the AR intervals.

The conventional Wald intervals improve more rapidly as a increases than do the AR intervals. When $a = 8$, and even more so when $a = 16$, the former have much better properties than the latter. The median length of the Wald intervals is just slightly smaller than the dispersion of the LIML estimates, while the median length of the AR intervals is much greater. Moreover, the distribution of the lengths of the Wald intervals is much tighter than that of the AR intervals. For $a = 16$, even the 0.99 quantile of the former is always smaller than the median length of the latter.

These results suggest that one would never want to use an AR confidence set when the instruments are reasonably strong. Even when they exist, AR intervals are much less informative than Wald ones. They do not have correct coverage conditional on

being bounded and non-empty. Moreover, they do not provide reliable information about the dispersion of $\hat{\beta}_{\text{LIML}}$; they can be much too long or much too short.

The results in Figure 4 complement those of Mikusheva (2010), who compares the AR confidence set with ones obtained by inverting the conditional likelihood ratio (CLR) test of Moreira (2003) and the LM test of Kleibergen (2002) and Moreira (2001), but not with conventional Wald tests. She finds that the AR interval tends to be somewhat longer, in expectation, than the CLR interval but considerably shorter than the LM interval based on the test of Kleibergen (2002).

6. Which Confidence Sets Should We Use?

The goal of this paper is simply to study the properties of AR confidence sets, not to settle the more difficult problem of which confidence set(s) to use when making inferences about β in equation (14) when the instruments are not strong. Davidson and MacKinnon (2013), a companion paper, tackles the latter problem.

The results of Mikusheva (2010) suggest that inverting the CLR test can work very well (and that inverting the LM test generally works poorly) and discusses how to invert the CLR test without using simulation. Davidson and MacKinnon (2013) proposes an explicit algorithm for inverting asymptotic CLR tests and finds that, in large samples, CLR confidence sets perform extremely well (unconditionally), even when the instruments are very weak.

Dufour’s (1997) result that any confidence set with correct coverage must be unbounded with positive probability in the neighborhood of a point at which the parameter of interest is not identified has consequences for any such possibly unbounded confidence set. The coverage probabilities conditional on the set being bounded, or on being unbounded, are not in general equal to the unconditional coverage probability. This fact is not an argument in favor of trying to use conditional coverage probabilities; rather, it underscores the undesirability of making inferences on nearly unidentified parameters.

Table 1 compares the behavior of AR confidence sets and ones obtained by inverting the CLR test for our reference case with $n = 400$ and $l - k = 7$, and for various values of a and ρ . It shows the percentages of the time when 95% confidence sets are empty, the whole real line, the real line with a hole, or bounded. It also shows the coverage probabilities conditional on the confidence set being neither empty nor the whole real line.

Only the AR confidence sets can be empty, but both can be unbounded with weak instruments. Both types of set exhibit coverage conditional on being bounded that is very different from the unconditional coverage when there is a non-negligible probability of an unbounded set. Like the AR sets, the CLR sets have very satisfactory unconditional coverage, but their conditional coverage is often quite unsatisfactory. The CLR sets always have a higher probability of being bounded than do the AR sets, and of course they cannot be empty. When $a = 8$, the CLR sets are all bounded, at least in our simulations, so that conditional and unconditional coverage are the

same, and very close to the nominal level. But there are still empty AR sets, and so coverage for bounded sets exceeds the nominal level.

In our view, there exist no circumstances in which one should use an AR confidence set. The dependence on the value of the Sargan statistic means that, even when the set is a bounded interval, its length provides very unreliable information about the precision with which the parameter of interest has been estimated. Moreover, as the instruments become stronger, the AR interval continues to perform poorly; see Figure 4. Thus the real defect of AR confidence sets is that, even when the instruments are strong, so that unbounded sets do not occur, the length of the sets can lead one to believe that the parameter of interest has been estimated either more or less precisely than is in fact the case.

CLR confidence sets certainly have merit and may well be worth using in practice, at least when the sample size is not too small. However, in addition to sometimes having seriously unsatisfactory conditional coverage, they have two practical disadvantages. First, they cannot readily be extended to handle two or more right-hand-side endogenous variables; see Mikusheva (2010). Second, because they are based on the LR statistic, they cannot easily deal with heteroskedasticity of unknown form.

Normally, higher power of an underlying test is reflected in shorter confidence intervals. Since the length of a bounded AR confidence set depends on the value of the Sargan statistic, test power must also depend on it. Without conditioning on the Sargan statistic, the relationship between the power of the AR test and that of other tests is complicated; see Davidson and MacKinnon (2008). It is shown there that the CLR test is, except for a few configurations of a and ρ , at least as powerful as the AR test. That is not the case for the LM test of Kleibergen (2002) and Moreira (2001), which is often much less powerful.

In contrast to CLR confidence sets, confidence intervals based on Wald tests can readily handle any number of endogenous variables and can easily be modified to allow for heteroskedasticity of unknown form. In our view, this type of interval has far more merit than it is generally given credit for. In Davidson and MacKinnon (2008), we proposed a procedure for bootstrapping t tests on β called the restricted efficient, or RE, bootstrap. In Davidson and MacKinnon (2010), we proposed a wild bootstrap variant of this procedure called the wild restricted efficient, or WRE, bootstrap that allows for heteroskedasticity of unknown form. Both procedures seem to work extraordinarily well, very much better than the pairs bootstrap or semiparametric bootstraps that do not impose restrictions, except when the instruments are extremely weak.

It is conceptually straightforward to invert t tests based on either $\hat{\beta}_{\text{LIML}}$ or $\hat{\beta}_{\text{IV}}$ that have been bootstrapped using either the RE or WRE bootstraps. The idea is simply to locate the ends of the interval at the points where the bootstrap P value is equal to α . This procedure can be computationally intensive, however. The problem is that, since the bootstrap DGP must impose a restriction on β , it is necessary to generate a different set of bootstrap samples for every candidate value of the upper

and lower limits of the confidence interval. If the interval has a hole, which is possible, it is also necessary to generate a set of bootstrap samples for every candidate value of the limits of the hole. Thus forming one confidence set can involve generating a great many bootstrap samples.

In Davidson and MacKinnon (2013), we present some simulation results for confidence intervals based on t statistics and the RE bootstrap, and we find that they generally work quite well, especially the ones based on $\hat{\beta}_{\text{LIML}}$. In large samples, they perform almost as well as asymptotic CLR intervals, provided the instruments are not very weak, and in small samples they perform much better.

7. Conclusion

It seems natural to make inferences about a parameter or parameters by inverting an exact test, such as the Anderson-Rubin test, because the resulting confidence set necessarily has correct coverage unconditionally. However, we argue that this is a very bad idea whenever the test has more degrees of freedom than there are parameters of interest. By considering two special cases, namely, inverting an F test and inverting an AR test for a scalar parameter, we show that the resulting confidence set provides very little useful information about the parameter of interest. It may be empty, extremely short, or excessively long. In the case of the AR confidence set, it may also be unbounded, although that is a problem shared by all confidence sets with good unconditional coverage when the instruments are weak.

The basic problem was explained in Section 2 in the context of inverting an F test to obtain a confidence set for a single parameter in a linear regression model. The problem arises because the confidence set depends not only on what the data tell us about that parameter but also on what they tell us about a number of additional restrictions. When those restrictions are moderately incompatible with the data, the confidence set will be a misleadingly short interval. When they are very incompatible with the data, it will not exist at all.

As we saw in Section 4, exactly the same problem arises in the context of the AR test. The additional restrictions are now the overidentifying restrictions that may be tested using a Sargan test. In this case, because the Sargan test is not exact, our results are necessarily asymptotic. When the Sargan statistic is moderately large, the AR confidence set is a misleadingly short interval. When it is very large, the AR confidence set does not exist. The simulation results in Section 5 provide strong support for our theoretical results and show that AR confidence sets can be very misleading even when the instruments are strong.

We emphasize that, although our analysis has treated only two special cases in detail, the conclusion about the misleading nature of confidence sets based on inverting tests with extra degrees of freedom over and beyond those needed for inference on the number of parameters of interest is quite general. In particular, when dealing with confidence sets for more than one parameter, the problem persists, although it may

well manifest itself in different ways, such as by giving rise to a confidence set with holes, like a torus.

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.
- 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 (2013), “Bootstrap confidence sets with weak instruments”, *Econometric Reviews*, forthcoming.
- Dufour, J.-M. (1997), “Some impossibility theorems in econometrics with applications to structural and dynamic models”, *Econometrica*, 65, 1365–1388.
- Dufour, J.-M., and M. Taamouti (2005), “Projection-based statistical inference in linear structural models with possibly weak instruments”, *Econometrica*, 73, 1351–1365.
- Forchini, G., and G. Hillier (2003), “Conditional inference for possibly unidentified structural equations”, *Econometric Theory*, 19, 707–743.
- Gleser, L., and J. Hwang (1987), “The non-existence of $100(1-\alpha)\%$ 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. (2009), “Tests with correct size when instruments can be arbitrarily weak”, *Journal of Econometrics*, 152, 131–140.
- 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.

- Sargan, J. D. (1958), “The estimation of economic relationships using instrumental variables”, *Econometrica*, 26, 393–415.
- 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.
- Stock, J. H., and M. Yogo (2005), “Testing for weak instruments in linear IV regression”, in D. W. K. Andrews and J. H. Stock (eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, 80–108. Cambridge: Cambridge University Press.
- 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.

Table 1

		\mathbb{R}	\emptyset	hole	bounded	coverage	hole	bounded
$a = 1$	$\rho = 0.10$	78.46	0.11	12.64	8.79	0.950	0.643	0.959
		82.06	—	8.09	9.85	0.942	0.426	0.880
	$\rho = 0.50$	76.66	0.13	14.37	8.84	0.950	0.726	0.892
		80.64	—	9.43	9.93	0.943	0.588	0.814
	$\rho = 0.95$	28.05	1.02	63.13	7.80	0.950	0.977	0.673
		33.66	—	53.36	12.98	0.944	0.957	0.746
$a = 2$	$\rho = 0.10$	60.98	0.34	14.29	24.39	0.950	0.737	0.962
		66.50	—	7.99	25.50	0.943	0.568	0.911
	$\rho = 0.50$	53.27	0.47	21.91	24.35	0.950	0.865	0.937
		59.33	—	13.27	27.40	0.943	0.757	0.909
	$\rho = 0.95$	0.04	2.49	75.09	22.38	0.950	0.986	0.935
		0.09	—	52.95	46.96	0.946	0.949	0.942
$a = 4$	$\rho = 0.10$	10.46	1.59	5.34	82.61	0.950	0.853	0.968
		13.98	—	2.75	83.26	0.945	0.771	0.941
	$\rho = 0.50$	3.83	1.98	12.09	82.11	0.950	0.944	0.971
		5.68	—	5.57	88.76	0.946	0.840	0.949
	$\rho = 0.95$	0.00	2.84	15.94	81.23	0.950	0.976	0.979
		0.00	—	2.89	97.11	0.947	0.465	0.961
$a = 8$	$\rho = 0.10$	0.00	2.68	0.00	97.32	0.950	N/A	0.976
		0.00	—	0.00	100.00	0.947	N/A	0.947
	$\rho = 0.50$	0.00	2.71	0.00	97.29	0.950	N/A	0.977
		0.00	—	0.00	100.00	0.947	N/A	0.947
	$\rho = 0.95$	0.00	2.84	0.00	97.16	0.950	N/A	0.978
		0.00	—	0.00	100.00	0.947	N/A	0.947

The middle block of columns gives the percentage for each type of confidence set: \mathbb{R} = whole real line, \emptyset = empty set, hole = real line with a hole, bounded = finite bounded interval. The rightmost block gives the unconditional coverage, followed by coverage conditional on a set that is the real line with a hole, and finally coverage conditional on a bounded interval. For each pair of values for a and ρ , results for the AR set are in the first line, and results for the CLR set are in the second line.

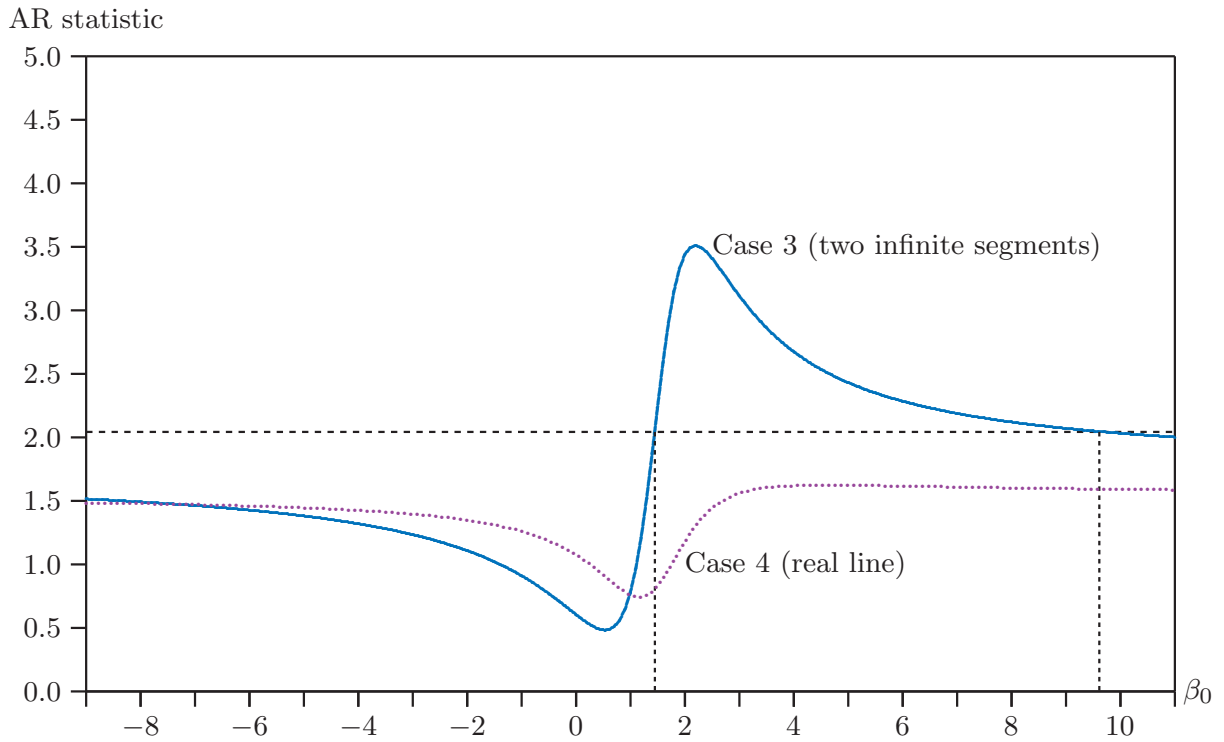
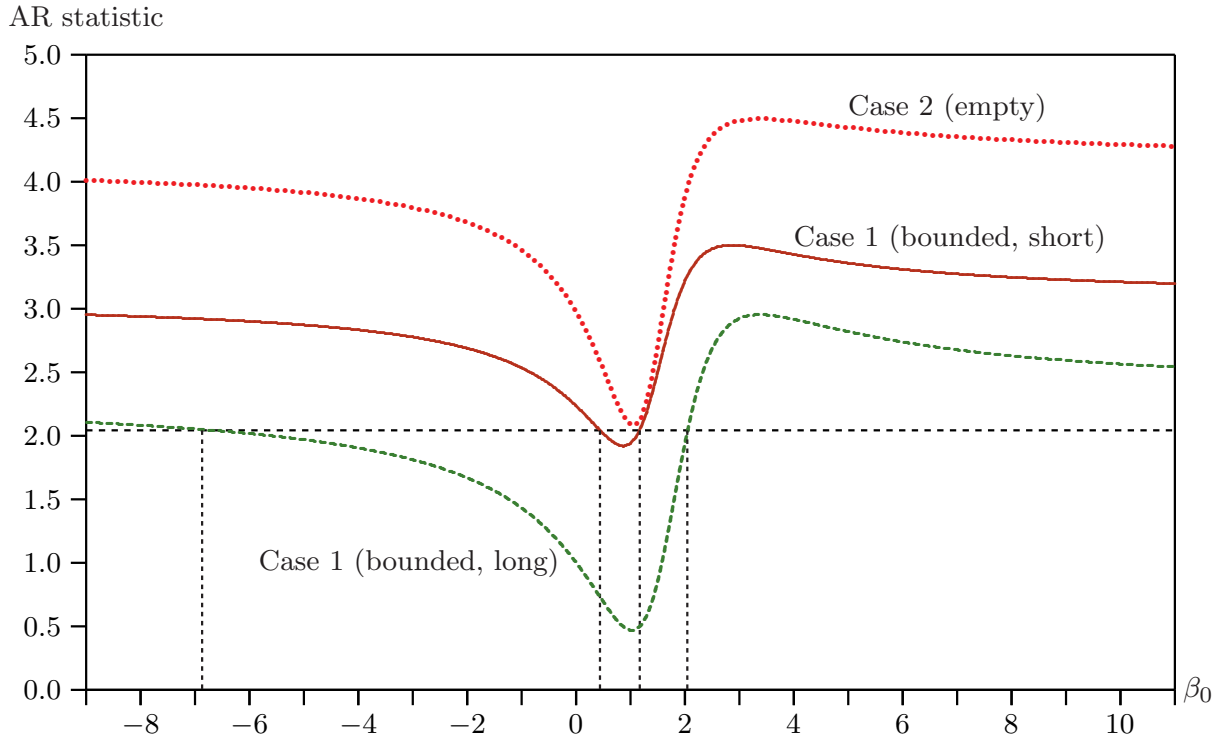


Figure 1. Five types of Anderson-Rubin confidence set

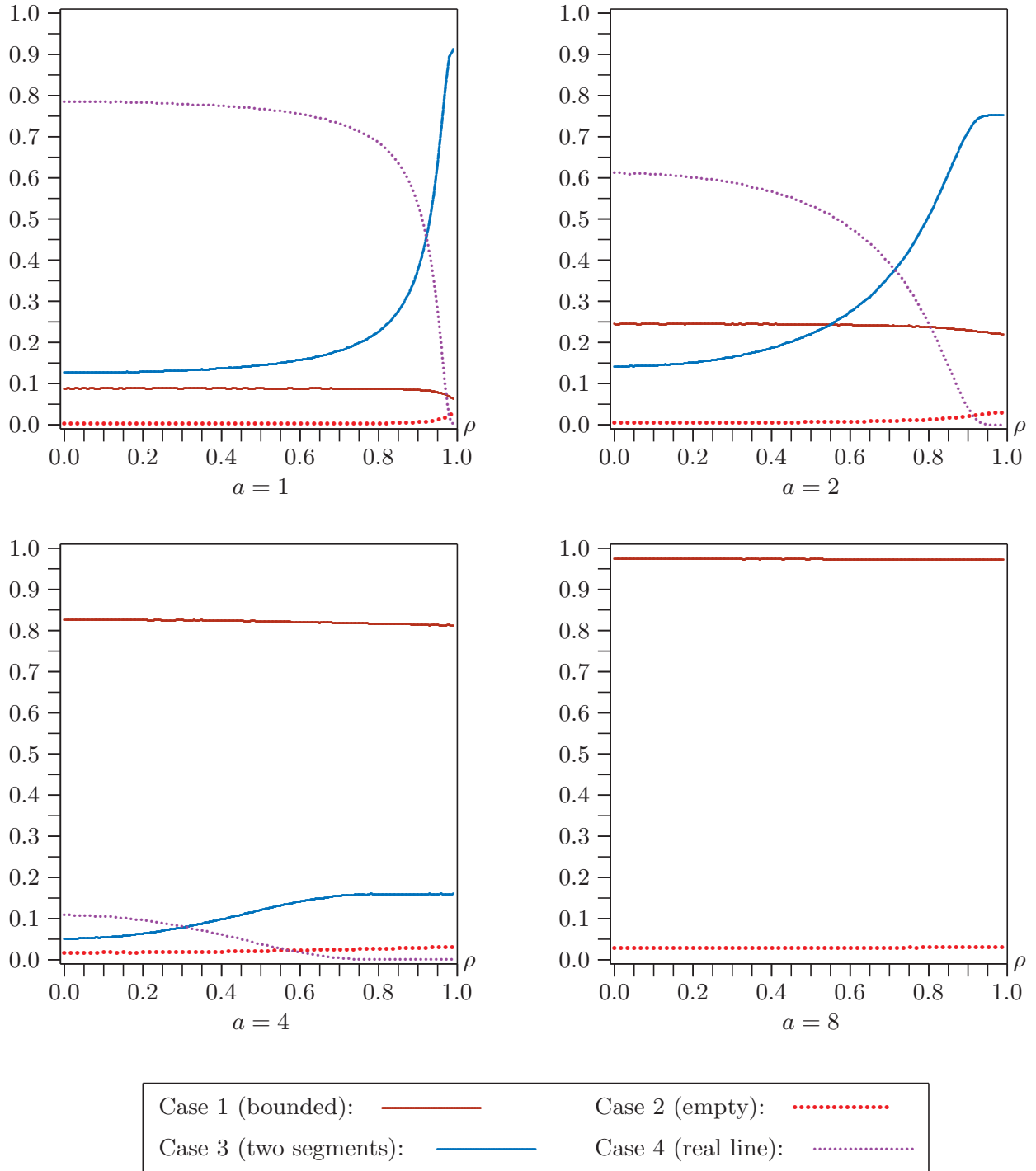


Figure 2. Frequencies of each type of confidence set as functions of ρ

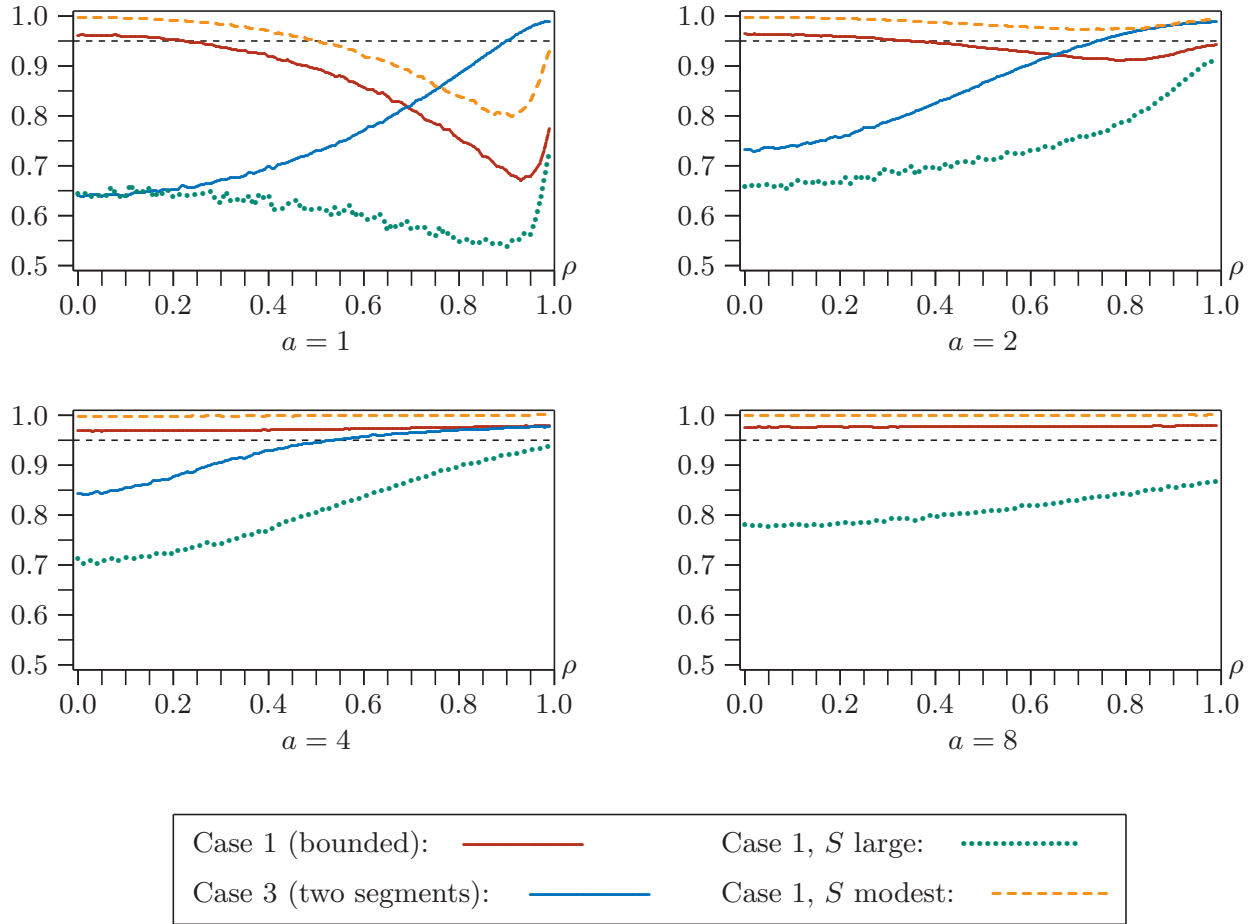


Figure 3. Coverage of each type of confidence set as functions of ρ

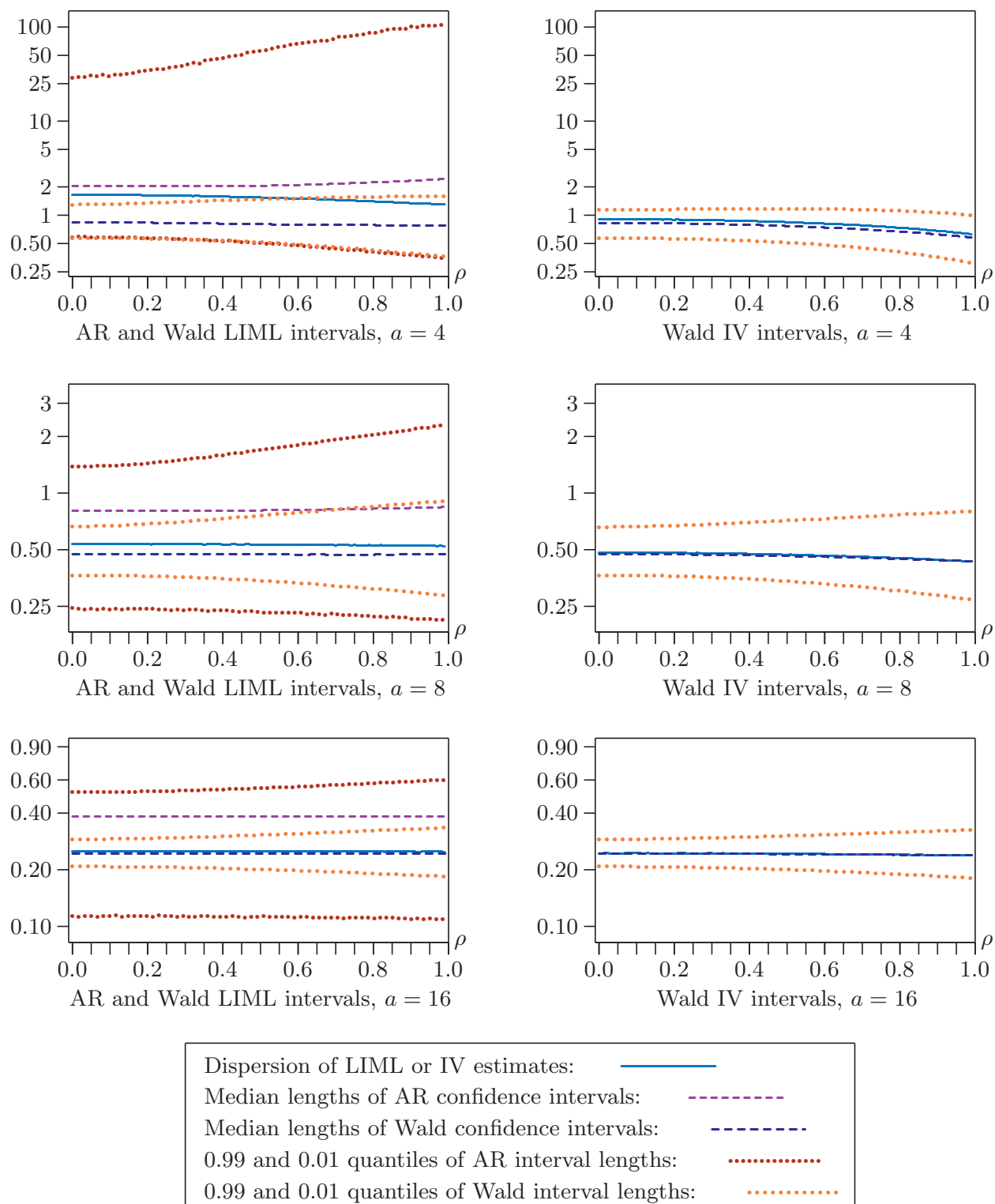


Figure 4. Dispersion of estimates and lengths of confidence intervals