Symmetry-based inference in an instrumental variable setting

Paul A. Bekker, Steve Lawford

▶ To cite this version:

HAL Id: hal-00501794
https://hal.science/hal-00501794

Submitted on 12 Jul 2010

Author's Accepted Manuscript

Symmetry-based inference in an instrumental variable setting

Paul A. Bekker, Steve Lawford

www.elsevier.com/locate/jeconom

Symmetry-based inference in an instrumental variable setting[*]

PAUL A. BEKKER AND STEVE LAWFORD[†]

Department of Economics, University of Groningen;

and CORE, Catholic University of Louvain,

and ECARES, Free University of Brussels

Abstract

We describe exact inference based on group-invariance assumptions that specify various forms of symmetry in the distribution of a disturbance vector in a general nonlinear model. It is shown that such mild assumptions can be equivalently formulated in terms of exact confidence sets for the parameters of the functional form. When applied to the linear model, this exact inference provides a unified approach to a variety of parametric and distribution-free tests. In particular, we consider exact instrumental variable inference, based on symmetry assumptions. The unboundedness of exact confidence sets is related to the power to reject a hypothesis of underidentification. In a multivariate instrumental variables context, generalizations of Anderson-Rubin confidence sets are considered.

1 Introduction

There is a large literature that tackles departures from standard asymptotics in an instrumental variables (IV) context. Recent theoretical contributions, with further references, can be found in Bekker (1994), Staiger and Stock (1997), Wang and Zivot (1998), Zivot et al. (1998), Donald and Newey (2001), Hahn and Hausman (2002), Kleibergen (2002, 2004) and Moreira (2001, 2003). In particular, Nelson and Startz (1990a, 1990b) considered the bimodality of the finite-sample distribution of the IV-estimator in the presence of a single instrument. Especially when the instrument is weak, in the sense of not being highly correlated with the regressor, when the degree of endogeneity is high, or when the number of observations is small, the asymptotic distribution may be a particularly poor approximation to the true distribution (additional details are provided by Woglom, 2001). An understanding of the empirical importance of weak instruments in econometrics stems from Angrist and Krueger's (1991) study of returns to education, and the discussion of their results by Bound et al. (1995). See also Imbens and Rosenbaum (2005), who examine permutation methods, and links with instrumental variables tests when instruments are weak, with an application to the Angrist and Krueger (1991) dataset. Good surveys of the weak instrument literature are Stock et al. (2002) and Hahn and Hausman (2003). In this paper, we consider a general distribution-free approach to exact inference, with particular emphasis on (weak) instrumental variables.

Instead of considering a specified parametric family of densities for the disturbances, we examine a variety of nonparametric assumptions based on three basic types of symmetry in the distribution of the disturbance vector. Such group-invariance assumptions can be very mild. (a) The exchangeability assumption, based

on the permutation group, is weaker than assuming that the disturbances are independent and identically distributed (iid). In fact, it also covers the case where sampling is performed without replacement. (b) Another group-invariance assumption, based on the reflection group, covers the heteroscedastic case, where the disturbances are assumed to be independent with symmetric distributions. This is weaker than the classical parametric assumption of iid Gaussian disturbances. (c) However, the assumption of iid Gaussianity can also be related to group invariance by considering the infinite group of rotations.

Furthermore, based on such assumptions, the paper presents exact inference on the parameters of the functional form. The inference is formulated in terms of exact confidence sets whose validity does not depend on the number of observations. So, contrary to asymptotic approximations, the approach may be applied locally, based on very few observations. Moreover, the inference is shown to be equivalent to the assumptions of the model if the group-invariance assumption is based on a finite group. In that case, the assumptions are necessary and sufficient for the inference to hold true. In particular, we consider such 'assumption-equivalent' inference in the linear model in an instrumental variable setting.

An important special case is given by linear regression, and as a first illustration consider the simple model:

$$y_i = \alpha + \beta x_i + u_i, \qquad i = 1, \ldots, n.$$

Let the elements y_i, x_i and u_i be collected in n-vectors y, x and u, respectively, and let P_j, $j = 1, \ldots, n! - 1$, represent all $n \times n$ permutation matrices that are different from the identity matrix I_n, so that $P_j x$ contains the elements of x in

a different order. Let $P_j x \neq x$, and consider points

$$c_j = \frac{x'(I_n - P_j)y}{x'(I_n - P_j)x}, \qquad j = 1, \ldots, n! - 1,$$

such that the real line is partitioned into $n!$ disjoint subsets B_k. It will be shown that the assumption that states that the elements of u are iid, and independent of x, is sufficient for exact inference that states that the probabilities for the $n!$ events $\beta \in B_k$ are all equal. Consequently, the sets B_k form elementary confidence sets for β. In fact, this inference is essentially equivalent to the assumption that says that the distribution of u is not affected by a reordering of its elements. A general formulation of this result, applicable to assumptions related to other groups of transformations, will be derived in the next section without reference to estimation or testing procedures.

This paper provides a unified approach to exact inference in the linear model under a variety of nonparametric assumptions. Many nonparametric statistical techniques, and early nonparametric inference in econometrics, can be recognized as special cases of our general framework, and some of these settings are briefly discussed. Descriptions of such distribution-free methods can be found in, inter alia, Hájek (1969), Lehmann (1975), Dawid (1988) and Maritz (1995). However, our main interest in this paper is in applications in an instrumental variables context.

For the general instrumental variable case, as considered here, inference becomes more complicated than it is for the linear regression. That is, inference on a single parameter remains assumption-equivalent, but the exact confidence regions need no longer be convex, symmetric or bounded. In particular, we study the possibility of unbounded confidence sets. The probability of unboundedness

is shown to be related to the power of the 'first-stage regression' to reject a hypothesis of under-identification.

Following the main result in Section 2, we focus on inference on a single parameter, and distinguish between two cases. First, in Section 3, monotonic inference is considered, which allows for one-sided inference, and where there are no complications due to unbounded confidence sets. Second, in Section 4, we assess inference based on a single instrument, which is nonmonotonic in general. We then discuss problems of interpretation that are related to exact confidence sets whose boundedness is not guaranteed. In Section 5, we explore joint confidence regions and inference based on more than one instrument. We describe nonparametric generalizations of Anderson-Rubin confidence sets and examine dynamic and nonlinear models. Section 6 concludes the paper.

2 Symmetry and exact inference

We consider nonparametric assumptions on the distribution of a disturbance vector. That is, instead of restricting the distribution to a parametric family, the distribution will be assumed to satisfy only a symmetry property. The mathematical expression of symmetry is invariance under a suitable group of transformations. First, we consider finite groups.

2.1 Symmetry assumptions based on finite groups

Based on elementary properties of finite groups, we begin by describing a result that can be formulated without using probability. Let the finite set $\mathcal{P} = \{P_0, P_1, \ldots, P_N\}$ of $n \times n$ matrices be closed under the formation of products and inverses, so that $\mathcal{P}$ is a finite transformation group of order $N + 1$. The identity matrix I_n is

an element, which will be denoted by P_0. Such a group defines an equivalence relation between n-vectors ε and ε^*, i.e. $\varepsilon \sim \varepsilon^*$ (mod $\mathcal{P}$) if $\varepsilon = P\varepsilon^*$ and $P \in \mathcal{P}$. An equivalence class, or 'orbit', will be denoted by $\mathcal{C}_\varepsilon = \{\varepsilon, P_1\varepsilon, \ldots, P_N\varepsilon\}$. Notice that $\mathcal{C}_\varepsilon = \mathcal{C}_{\varepsilon^*}$ if $\varepsilon \sim \varepsilon^*$ (mod $\mathcal{P}$). Examples of such finite groups are given by permutation matrices, where $P\varepsilon$ contains the elements of ε in a different order, and $N = n! - 1$; and by reflection matrices, which are diagonal with diagonal elements equal to either 1 or -1, and $N = 2^n - 1$. A combination, containing both permutations and reflections, is formed by the permutation-reflection group where $N = n!2^n - 1$. Subgroups will also be considered. For example, if the elements of ε are stratified, permutations within strata form a subgroup, which can be achieved by block-diagonal permutation matrices.[1]

Let $g(\varepsilon)$ be a scalar *inferential function*. Then, the range of $g(\cdot)$ with domain $\mathcal{C}_\varepsilon$, i.e.

$$\mathcal{R}(g(\varepsilon)) \equiv \{g(P\varepsilon) \mid P \in \mathcal{P}\},$$

is the *reference set* of $g(\varepsilon)$. From the group structure, $\mathcal{R}(g(P\varepsilon)) = \mathcal{R}(g(\varepsilon))$ if $P \in \mathcal{P}$. This property, and Condition 1, make it possible to formulate assumption-equivalent inference.

CONDITION 1: *All elements of the reference set $\mathcal{R}(g(\varepsilon^*))$ are different, i.e.* $g(P_i\varepsilon) \neq g(P_j\varepsilon)$ *for all* $i, j \in \{0, 1, \ldots, N\}$ *and* $i \neq j$.

Consequently, if Condition 1 is satisfied, there is a one-to-one correspondence

---

[1] We use "strata" in a standard statistical sense, see e.g. Imbens and Rosenbaum (2003, Section 3.1), for further details.

between the elements of $\mathcal{C}_{\varepsilon^*}$ and the elements of $\mathcal{R}(g(\varepsilon^*))$. If we define[2]

$$F(\varepsilon) \;\equiv\; \#\{P \mid g(\varepsilon) < g(P\varepsilon), P \in \mathcal{P}\},$$

with range $\mathcal{U} = \{0, 1, \ldots, N\}$, then the mapping $F : \mathcal{C}_{\varepsilon^*} \to \mathcal{U}$ is one-to-one. Conditional on $\varepsilon \in \mathcal{C}_{\varepsilon^*}$, where Condition 1 is satisfied, the following two statements are equivalent:

$$\varepsilon \;=\; P_i\varepsilon^*, \tag{1}$$

$$F(\varepsilon) \;=\; F(P_i\varepsilon^*). \tag{2}$$

The interpretation of this equivalence becomes clearer in an econometric context, where ε is a random disturbance vector. In particular, we consider disturbances specified as

$$\varepsilon = f(y, X; \beta_o),$$

where the functional form $f(y, X; \beta)$ is a known function of both observable variables (y, X) and a parameter vector β. We will use the short notation $F(\beta)$ instead of $F(f(y, X; \beta))$.

An assumed symmetry property of the distribution of ε amounts to a *group invariance assumption* which states that ε and $P\varepsilon$ have the same distribution if $P \in \mathcal{P}$. When conditioned on equivalence classes $\mathcal{C}_{\varepsilon^*}$, where Condition 1 is satisfied, such a group invariance assumption states that the conditional distribution of ε is uniform over $\mathcal{C}_{\varepsilon^*}$. From the equivalence of (1) and (2), we find that the group invariance assumption holds, conditional on $\mathcal{C}_{\varepsilon^*}$, if and only if the

---

[2]Similar to the generalized Bell-Pitman statistic, given by $1 - F(\varepsilon)(N + 1)^{-1}$, cf. Dawid (1988).

conditional distribution of $F(\beta_o)$ is uniform over $\mathcal{U}$. The latter amounts to exact assumption-equivalent inference on β_o. Our findings are summarized as follows.

PROPOSITION 1: *Conditional on an equivalence class, where Condition 1 is satisfied, a group invariance assumption stating that ε is uniformly distributed over $\mathcal{C}_{\varepsilon^*}$ is equivalent to inference on β_o that says that $F(\beta_o)$ is distributed uniformly over $\mathcal{U}$. If Condition 1 holds almost surely (a.s.), then a group invariance assumption implies that the marginal distribution of $F(\beta_o)$ is uniform over $\mathcal{U}$.*

If $F(\beta_o)$ is indeed uniformly distributed over $\mathcal{U}$, the function $F(\beta)$ can be used to construct exact (as opposed to approximate asymptotic) confidence sets for β_o:

$$\mathcal{C}(\mathcal{V}) \equiv \{\beta \mid F(\beta) \in \mathcal{V} \subset \mathcal{U}\}; \tag{3}$$

with size $\mathrm{Prob}(\beta_o \in \mathcal{C}(\mathcal{V})) = \#\mathcal{V}/\#\mathcal{U}$. If $\mathcal{V}$ consists of a single element, $\mathcal{C}(\mathcal{V})$ is called an *elementary confidence set.*

Due to computational problems, it may be difficult to implement this inference in practice if N is a large number. However, as is shown in the Appendix, exact inference can also be formulated based on a random sample of matrices $P \in \mathcal{P}$. The function $F(\beta)$, based on this random sample, can be computed for a grid of values for β. Confidence sets can then be constructed by trial and error, i.e. $\beta \in \mathcal{C}(\mathcal{V})$ if $F(\beta) \in \mathcal{V}$. In some cases, inference on a single element of the parameter vector can also be formulated by computing confidence limits directly. These are the bounds of one-dimensional elementary confidence sets.

2.2 Symmetry based on the infinite orthonormal group

The set $\mathcal{P}$ of $n \times n$ orthonormal matrices P, with $P'P = PP' = I_n$, forms an infinite group. Nevertheless, an infinite group-invariance assumption implies exact inference conditional on an equivalence class, and therefore also unconditionally. The group-invariance assumption now says that the vector ε can be rotated without affecting its distribution. So, ε is assumed to have a spherically symmetric distribution[3]: it depends only on $\varepsilon'\varepsilon$. A rotational-invariance assumption implies permutation-reflection invariance. If the elements of ε are also assumed to be independent, rotational invariance amounts to Gaussianity.

Let $\{P_1, \ldots, P_N\}$ be a random sample taken from $\mathcal{P}$ and let $P_0 = I_n$. That is, for any ε, the elements of $\{P_1\varepsilon, \ldots, P_N\varepsilon\}$ are assumed to be independently uniformly distributed over the surface of the sphere with radius $(\varepsilon'\varepsilon)^{1/2}$. Thus, ε and $P_1\varepsilon$, say, have identical distributions. If we further condition on an equivalence class $\varepsilon \in \mathcal{C}_{\varepsilon^*} = \{P\varepsilon^* \mid P \in \mathcal{P}\}$, then ε and $P_1\varepsilon$ are also independent. Consequently, conditional on $\mathcal{C}_{\varepsilon^*}$, the random vectors $P_i\varepsilon$, and $g(P_i\varepsilon)$, $i = 0, 1, \ldots, N$ will be iid. Hence, $F(\beta_o) \sim U(\mathcal{U})$, where

$$F(\beta_o) = \#\{i \mid g(\varepsilon) < g(P_i\varepsilon), i = 1, \ldots, N\}, \tag{4}$$

which holds conditionally and unconditionally. We may also consider infinite subgroups where the rotational symmetry applies to linear subspaces.

---

[3]The argument can easily be generalized to elliptical distributions.

3 Monotonic inference and some applications

Exact inference as formulated in Proposition 1 is not unique. Each choice of inferential function $g(\cdot)$ that satisfies Condition 1, a.s., provides exact inference. Such inferences can be distinguished based on aspects of the shape of the random function $F(\beta)$. That is to say, $F(\beta)$ should ideally allow relevant confidence sets to be bounded, convex and non-empty a.s. For this reason, we pay particular attention to inference on a single parameter based on linear inferential functions. We then distinguish between monotonic inference, where $F(\beta)$ is a monotonic function with range $\mathcal{U}$, and nonmonotonic inference. Section 5 also discusses nonlinear inferential functions.

3.1 Monotonic inference

Inference based on $F(\beta_o) \sim U(\mathcal{U})$ is monotonic if β is a scalar and $F(\beta)$ is a monotonically increasing function ranging from $F(-\infty) = 0$ to $F(\infty) = N$. Then, 'assumption-equivalent' inference amounts to the formulation of an ordered collection of $N+1$ elementary convex sets that partition the real line. Using notation (3), these elementary confidence sets can be denoted $\mathcal{C}(\{i\})$, $i = 0, 1, \ldots, N$. Their coverage probabilities are all equal to $(N+1)^{-1}$. The following is a simple illustration of *monotonic inference*.[4]

Consider the example used in the Introduction:

$$\varepsilon = y - \beta_o x = \alpha \iota_n + u, \tag{5}$$

---

[4]Such inference is related to *distributional inference* as discussed by Kroese and Schaafsma (1998), who consider optimality in terms of proper loss functions based on the difference between random distribution functions $G(\beta)$, where $G(\beta_o) \sim U([0,1])$, and the indicator function of $[\beta_o, \infty)$.

where ι_n is an n-vector of ones. The distribution of ε, conditional on x, is assumed to be invariant under permutations of its elements. Therefore, ε and $P\varepsilon$ have the same distribution if $P \in \mathcal{P}$, where $\mathcal{P}$ is the permutation group of $n \times n$ matrices. Let $g(\varepsilon) = x'\varepsilon$ and assume that Condition 1 is satisfied a.s. This implies, for $i = 1, \ldots, n! - 1$, that $P_i x \neq x$ and, since P_i is orthogonal,

$$x'(I_n - P_i)x = \frac{1}{2}x'\left[(I_n - P_i) + (I_n - P_i)'\right]x = \frac{1}{2}x'(I_n - P_i)(I_n - P_i)'x > 0. \tag{6}$$

Consequently, assumption equivalent inference amounts to $F(\beta_o) \sim U(\mathcal{U})$, where[5]

$$F(\beta) = \#\left\{i \mid \frac{x'(I_n - P_i)y}{x'(I_n - P_i)x} < \beta, i = 1, \ldots, N\right\}, \tag{7}$$

and $N = n! - 1$. Here we need not construct confidence sets by trial and error. Instead, we can compute confidence limits c_i directly:

$$c_i = \frac{x'(I_n - P_i)y}{x'(I_n - P_i)x}, \ i = 1, \ldots, n! - 1. \tag{8}$$

We may also consider a random sample of these limits, as described in the Appendix.[6]

We find for monotonic inference that elementary confidence sets are non-empty and convex, and that only $\mathcal{C}(\{0\})$ and $\mathcal{C}(\{N\})$ are unbounded. Due to the

---

[5]Note that $g(\varepsilon) < g(P\varepsilon)$ as $x'\varepsilon < x'P\varepsilon$ as $x'(I_n - P_i)y < x'(I_n - P_i)x\beta$, and (7) follows from property (6). Here, $g(\varepsilon) - g(P\varepsilon) = x'(I_n - P)u$, and so permutation based inference for β is independent of α.

[6]The inference described here may perhaps be seen as similar to Monte Carlo methods such as the bootstrap, which is a technique for estimating the finite-sample distribution of a statistic, or a feature thereof, by data resampling. Bootstrap estimation provides an *approximation* to exact finite-sample methods (e.g. approximation of a distribution function), and is usually justified by *asymptotic* arguments under suitably general conditions. By contrast, the methods in this paper lead to *exact* inference, whether based on the complete group $\mathcal{P}$, or a random sample from $\mathcal{P}$. See Robinson (1987) for comparison of several bootstrap confidence intervals as approximations to exact intervals derived from inversion of exact permutation tests.

ordering of the elementary confidence sets, one-sided inference can be formulated based on confidence sets $\mathcal{C}(\mathcal{V})$, or their complements, where $\mathcal{V} = \{j, \ldots, N\}$. Two-sided inference can be based on confidence sets given by

$$\mathcal{S}_j = \mathcal{C}(\mathcal{V}_j); \tag{9}$$

$$\mathcal{V}_j = \{j, \ldots, N - j\}, \ 0 \le j \le N/2.$$

So, we find that

$$\mathrm{Prob}(\beta_o \in \mathcal{S}_j) = 1 - \alpha_j, \tag{10}$$

$$\alpha_j = 2j(N + 1)^{-1} = 1 - \#\mathcal{V}_j/\#\mathcal{U}.$$

For monotonic inference, the sets $\mathcal{S}_j$ are convex, non-empty and also bounded for $j > 0$. Moreover, they are *symmetric*, which is of some importance when interpreting the sets $\mathcal{S}_j$. In particular, the median of the N confidence limits – with some tolerance if N is even – has a distribution whose median coincides with β_o. Therefore, similar to the Hodges-Lehmann estimator (cf. Hodges and Lehmann, 1963; or Lehmann, 1975, p. 82), this median is a *median unbiased estimator* of β_o. In other words, $\mathcal{S}_j$ has coverage probability $1 - \alpha_j$ for scalar β_o, and a median unbiased estimator $\widehat{\beta}_{\mathrm{med}}$, in the *scalar monotonic* case, is given by $\widehat{\beta}_{\mathrm{med}} = \lim_{\alpha_j \to 1} \mathcal{S}_j$. See Andrews (1993) and Andrews and Phillips (1987) for some discussion of median-unbiasedness in econometrics, and Lehmann (1997, Section 3.5) for an introduction.

Graphically, the inference can be represented by an 'F-plot', where $F(\beta)$ is

plotted against β. An alternative two-sided representation is given by a P-plot, or 'confidence pyramid', as used by Gabriel and Hall (1983) and Tritchler (1984), where the two-sided confidence regions are stacked on top of each other. That is, consider values $P(\beta)$:[7]

$$P(\beta) = \max\{\alpha_j \mid \beta \in \mathcal{S}_j, j = 0, \ldots, N\} = \frac{N}{N+1}\left(1 - \left|1 - \frac{2F(\beta)}{N}\right|\right). \quad (11)$$

Then, $H_o : \beta = \beta_o$ would be rejected at significance level α_j, (i.e. $\beta_o \notin \mathcal{S}_j$), if and only if $\alpha_j > P(\beta_o)$. So, $P(\beta)$ provides a representation of the sets $\mathcal{S}_j$:

$$\mathcal{S}_j = \{\beta \mid P(\beta) \geq \alpha_j\}. \quad (12)$$

As an example of monotonic inference based on model (5), where $F(\beta)$ takes the form (7), consider iid data generated as follows: $\varepsilon_i = (w_1 + w_2^2 + 3)_i^2$ and $x_i = \mid \log \mid \frac{w_3}{w_4} \mid \mid$, where for each $i = 1, \ldots, 10$, independently, $w \sim \mathcal{N}(0, I_4)$; $\beta_o = 1$. Figure 1 gives a P-plot based on 1000 randomly chosen permutations, and a median-unbiased estimate $\widehat{\beta}_{\mathrm{med}} > \beta_o$.

[Figure 1 about here.]

The assumption-equivalent inference developed here is exact, for all sample sizes n, irrespective of the number of random samples taken from the (permutation or other) group $\mathcal{P}$. Clearly, a "reasonable" subsample size will depend upon the computing resources available, and the context, and will likely be case-dependent. As an illustration, we performed a small study, using the model following equation (12). We construct confidence sets $\mathcal{S}_j$ for $n = \{10, 25\}$, using $N = 10,000$ random $n \times n$ permutation matrices, where $\alpha_j \in \{0.05, 0.10, 1.00\}$.

---

[7]An alternative would be to use p-values: $\min\{\alpha_j \mid \beta_j \notin \mathcal{S}_j, j = 1, \ldots, N\}$.

The procedure was repeated 200 times, and we recorded the empirical mean and standard deviation of the lower and upper bounds of the confidence sets with 95% and 90% coverage, and the median unbiased estimate $\widehat{\beta}_{\mathrm{med}}$. We found that reasonable accuracy is achieved (in terms of small variability across samplings from $\mathcal{P}$), even for small samples, and especially for the median-unbiased estimate. For instance, the 90% interval for $n = 25$ is given by $(-3.34, 7.06)$, with estimated standard errors 0.067 and 0.061 on the bounds. In a practical application, N could be increased considerably. For the same model, we constructed P-plots, using different subsamples from $\mathcal{P}$, where $N \in \{10, 100, 500, 1000, 2500, 10000\}$. It appears that N as small as 1000 gives quite accurate inference in this example (although *all* display exact inference).

3.2   Some special cases

Exact inference based on confidence limits (8) was derived by Gabriel and Hall (1983) and Maritz (1995) by inversion of permutation tests.[8] Based on asymptotic conditions, the difference between this nonparametric inference and Gaussian inference frequently becomes negligible even for modest values of n. However, if the regression is applied locally, e.g. as a building-block for nonparametric regression, based on very few observations, then differences may be seen. Another case arises when asymptotic conditions are not well satisfied, such as when two samples are compared, one of which is small (for example, Wilcoxon's two-sample test, or similarly the Mann-Whitney test, easily fit within the present framework). We now give some examples to illustrate how the group-invariance framework can provide a unified treatment of various different problems.

---

[8]In an experimental context, given similar assumptions with respect to x instead of ε, exact inference of the form (10) would still be valid.

EXAMPLE 1 (Permutation):

The question is whether the permutation argument can also be applied in a general multiple regression context:

$$y = x\beta + X_2\gamma + u = x\beta + \varepsilon, \tag{13}$$

where the elements of u are exchangeable. In that case, the matrices P_i should satisfy the additional condition $(I_n - P_i)X_2 = 0$, in order to filter out the effect of the variables X_2. In other words, the transformations P_i should affect the vector x, i.e. $P_i x \neq x$, but they should leave the matrix X_2 unaffected, i.e. $P_i X_2 = X_2$. In order to achieve this, the permutations should be performed between observations where the variables in X_2 are constant. In particular, if X_2 consists of dummy variables, defining strata, this nonparametric approach is applicable by using a subgroup consisting of block-diagonal permutation matrices. In case X_2 does not consist of dummy variables, then joint confidence sets for the parameters in β and γ might be formulated as described in Section 5.

EXAMPLE 2 (Reflection):

Another well-known application relates to the central location of symmetric distributions. Let $y = \beta \iota_n + \varepsilon$ (in fact, it is only necessary that the constants multiplying β are different from zero), and let $\mathcal{P}$ be the reflection group. From the independence and the symmetry of the distributions, ε and $P\varepsilon$ have the same distribution if $P \in \mathcal{P}$. So, if we use $g(\varepsilon) = \iota_n' \varepsilon$ as an inferential function, and

assume that Condition 1 is satisfied, we find using similar techniques to above that $F(\beta_0)$ will be uniformly distributed over $\mathcal{U}$, where $N = 2^n - 1$ and

$$
F(\beta_0) \quad = \quad \# \left\{ i \mid \frac{\iota_n'(I_n - P_i)y}{\iota_n'(I_n - P_i)\iota_n} < \beta, i = 1, \ldots, 2^n - 1 \right\}.
$$

This holds as $\iota_n'(I_n - P_i)\iota_n > 0$, as $n = \iota_n'\iota_n > \iota_n'P_i\iota_n$. Since $P_i \in \mathcal{P}$ (reflection group), $i = 1, \ldots, N$, is diagonal, $\iota_n'P_i\iota_n = \text{trace}(P_i)$. Noting that the diagonal elements of $P_i \neq I_n$ are either 1 or -1, it follows directly that $\text{trace}(P_i) < n$. The set of confidence limits is now given by the means of all $2^n - 1$ subsets of $\{y_1, \ldots, y_n\}$. Maritz (1995) derives these limits based on a permutation argument used by R.A. Fisher. A discussion of the Fisher randomization test is given in Basu (1980).

EXAMPLE 3 (Rotation):

Perhaps the most well-known example is given by the assumption of rotational invariance in the context of linear regression. Consider the multiple regression equation (13). Now assume that the conditional distribution of u, given $(x, X_2) : n \times (k + 1)$, is spherical. Then, the distribution of ε is not affected by rotations over a linear subspace. That is, let $P_{X_2} = X_2(X_2'X_2)^{-1}X_2'$, and let L be an $n \times (n - k)$ orthonormal complement of X_2, i.e. $L'X_2 = 0$ and $L'L = I_{n-k}$. Now, consider the set

$$
\mathcal{P} \quad = \quad \{P \mid P = P_{X_2} + L\tilde{P}L', \tilde{P}'\tilde{P} = I_{n-k}\},
$$

where $\tilde{P}$ is $(n - k) \times (n - k)$. Clearly, $\mathcal{P}$ forms a group, and $PX_2 = X_2$, so that

$P\varepsilon = X_2\gamma + Pu$, for any $P \in \mathcal{P}$. Consequently, ε satisfies the group-invariance assumption $\varepsilon \sim P\varepsilon$ for $P \in \mathcal{P}$. As a result, $F(\beta_o)$ as given in (4) will be uniformly distributed over $\mathcal{U}$. If we use $g(\varepsilon) = x'\varepsilon$ as an inferential function, we find confidence limits similar to those found for the other group invariance assumptions. In fact, the inference converges to classical inference based on the t_{n-k-1}-distribution as $N \to \infty$, which agrees with Efron (1969).

PROPOSITION 2 (Bekker, 2002): *Let $P = P_{X_2} + L\tilde{P}L'$, where, for any $v \in \mathrm{R}^{n-k}$, $\tilde{P}v$ is distributed uniformly over the surface of the sphere with radius $(v'v)^{\frac{1}{2}}$. Then, for any $a \in \mathrm{R}$,*

$$\mathrm{Prob}\left(a < \frac{x'(I_n - P)y}{x'(I_n - P)x} \mid x, y \right) \;\; = \;\; \mathrm{Prob}(a < \hat{\beta} + se(\hat{\beta})t_{n-k-1} \mid x, y),$$

where t_{n-k-1} is distributed as Student's t with $n - k - 1$ degrees of freedom; and $\hat{\beta}$ and $se(\hat{\beta})$ are the ordinary least squares estimator and its standard error, respectively.

Therefore, traditional exact inference, based on Gaussianity, fits within the present framework. Bekker (2002) also uses the present approach to describe 'optimal' exact inference – based on a minimum variance argument – in the case of groupwise heteroscedasticity, where homoscedasticity is restricted to strata.[9] For such a (feasible) WLS context, inference based on maximum likelihood is only approximate.

---

[9]Bekker (2002) explores a special case (rotation invariance) of the present general group invariance framework, in a single parameter linear model, given groupwise (hence, partial rotation invariance) heteroscedastic Gaussian innovations.

3.3 Measurement errors and monotonic inference

Another special case of monotonic inference is related to Wald's (1940) problem of fitting straight lines if both variables (y and x in (5)) are subject to error. Wald's approach was based on the assumption that iid errors in x did not affect its ranking. Aigner et al. (1984, p. 1339) describe Wald's method as an instrumental variables technique with classification dummy variables as instruments. In fact, Theil (1950) provided exact confidence limits for β_o (cf. Lehmann, 1975, p. 312; Maritz, 1995) given by:

$$\frac{y_i - y_j}{x_i - x_j}, \ \ i < j, \ \ i, j = 1, \ldots, n. \tag{14}$$

However, the resulting $1 + \sum_{k=1}^{n-1} k = \frac{1}{2}n(n-1) + 1$ confidence intervals do not have equal coverage probabilities. Maritz (1995) gives a simple representation of the distribution and, in a different context, it has been described by Kendall (1938, 1975) and Mann (1945). Here we can show there is another part to the story.

The assumption that the errors do not affect the rank numbers of x can be used to define an inferential function $g(\varepsilon) = r'\varepsilon$, where r contains the rank numbers of x. This ranking can be considered as an instrument. Due to the errors, ε and x are correlated. However, conditional on r, ε satisfies the exchangeability assumption. Furthermore, since all elements of x are different, Condition 1 is satisfied, and for all permutation matrices $P \neq I_n$ we find

$r'(I_n - P_i)x > 0$. Consequently, $F(\beta_o) \sim U(\mathcal{U})$, where

$$
\begin{aligned}
F(\beta) &= \#\{i \mid g(\varepsilon) < g(P_i \varepsilon), i = 1, \ldots, n! - 1\} \\
&= \# \left\{ i \mid \frac{r'(I_n - P_i)y}{r'(I_n - P_i)x} < \beta, i = 1, \ldots, n! - 1 \right\}.
\end{aligned}
$$

Considering the subset of confidence limits generated by simple permutation matrices, where only two elements are permuted, amounts to Theil's inference with confidence limits as in (14). However, this does not produce a uniform distribution because simple permutations do not form a group. Consideration of all possible permutations results in the monotonic, 'assumption-equivalent' inference presented here, i.e. $F(\beta_o) \sim U(\mathcal{U})$, and is more precise since it extends Theil's original results and produces additional exact confidence limits based on the same assumption.

4 Inference based on a single instrument

We now discuss assumption-equivalent inference based on a single instrument z, where ε satisfies a group-invariance assumption conditional on z, and the inferential function is $g(\varepsilon) = z'\varepsilon = z'(y - x\beta_o)$. First, we consider monotonic inference. Subsequently, nonmonotonic inference and the relation between identification and bounded confidence sets will be considered, as well as the coverage probability conditional on their boundedness.

4.1 Monotonic instruments

The instrument z is said to be *monotonic* with respect to x and the group $\mathcal{P}$ if $z'(I_n - P)x > 0$ a.s. for $P \in \mathcal{P}$, and $P \neq I_n$. In that case, inference based on

$F(\beta_o) \sim U(\mathcal{U})$, where

$$F(\beta) = \# \left\{ i \mid \frac{z'(I_n - P_i)y}{z'(I_n - P_i)x} < \beta, i = 1, \ldots, N \right\}, \tag{15}$$

will be monotonic. Note that, for the permutation group, the rank numbers of the elements of z and x should be the same. An example was given in the previous section, related to Theil's inference, where the rank numbers themselves formed the instrument. For the reflection group, the signs of the elements of z should be the same as those of x. For the permutation-reflection group, the rank numbers and their signs should be the same. Finally, the rotation group requires a monotonic instrument to be a scalar multiple of x.[10] These requirements may seem to be quite restrictive. However, the restrictions can be relaxed by considering subgroups. For instance, for stratified data, where a permutation-invariance assumption applies to the strata, a monotonic instrument should have the same ordering as x only within strata. Exact inference based on monotonic instruments with respect to a stratified rotational-invariance assumption is described in Bekker (2002).

---

[10] For instance, $z'(I_n - P_i)x > 0$ for $P_i \in \mathcal{P}$ (reflections) follows directly from $I_n - P_i = \mathrm{diag}(c_1, \ldots, c_n)$, $(c)_j \in \{0, 2\}$, $c_j \neq 0$ for some j, whereupon $z'(I_n - P_i)x = \sum_j (c)_j (z)_j (x)_j > 0$ for all z, x, iff $(z)_j$ and $(x)_j$ have same sign, for all j. Moreover, for $P_i \in \mathcal{P}$ (rotations), a geometric argument may be used: $z'(I_n - P_i)x > 0$ as $z'x > z'P_ix$ as $|z||x|\cos\theta > |z||P_ix|\cos\phi$, where θ and ϕ are the angles between z and x, and z and P_ix respectively. Hence, $\cos\theta > \cos\phi$ for all θ, ϕ iff $\cos\theta = 1$ (and $\theta = 0$). Directly, $z = \lambda x$, λ scalar. The proof for $\mathcal{P}$ (permutations) is trivial.

4.2   Nonmonotonic instruments

In general, an instrument z need not be monotonic. The analysis is based on $F(\beta_o) \sim U(\mathcal{U})$, where instead of (15) we now have

$$
\begin{aligned}
F(\beta) \;\;=\;\; & \# \left\{ i \mid \frac{z'(I_n - P_i)y}{z'(I_n - P_i)x} < \beta, z'(I_n - P_i)x > 0 \right\} + \\
& \# \left\{ i \mid \frac{z'(I_n - P_i)y}{z'(I_n - P_i)x} > \beta, z'(I_n - P_i)x < 0 \right\} + \\
& \# \left\{ i \mid z'(I_n - P_i)y < 0, z'(I_n - P_i)x = 0 \right\}.
\end{aligned}
\tag{16}
$$

We only consider cases where $N^0 \equiv \#\{i \mid z'(I_n - P_i)x = 0\} = 0$. Then $F(\beta)$ consists of two terms, $F(\beta) = F^+(\beta) + F^-(\beta)$, say, where $F^+(\beta)$ and $F^-(\beta)$ are based on N^+ and N^- points, respectively, and $N^+ \equiv \#\{i \mid z'(I_n - P_i)x > 0\}$ and $N^- \equiv \#\{i \mid z'(I_n - P_i)x < 0\}$. Hence, $N = N^+ + N^-$ and both $F^+(\beta)$ and $F^-(\beta)$ are monotonic step functions increasing from 0 to N^+ and decreasing from N^- to 0, respectively. If $N^+ > 0$ and $N^- > 0$ with positive probability, the nice properties of monotonic inference described above are lost.

(i) The elementary confidence sets $\mathcal{C}(\{i\})$, $i = 0, \ldots, N$, will not form an ordered collection of convex sets. Consequently, there is no exact one-sided inference.

(ii) Some sets $\mathcal{C}(\{i\})$ may be empty.

(iii) The two-sided confidence sets $\mathcal{S}_j$, defined in (9), need not be symmetric.

(iv) The top of a P-plot, defined in (11), need not be a median unbiased estimator of β_o.

(v) The sets $\mathcal{S}_j$ may be unbounded for positive values of j with positive probability.

As a result of this nonmonotonicity, the interpretation of the exact confidence sets $\mathcal{S}_j$ is more complicated than in the monotonic case. Fortunately, the sets $\mathcal{S}_j$ cannot be empty if $N^0 = 0$. Furthermore, the negative effects of non-monotonicity need not be severe, which holds in particular for large samples. Let the *degree of monotonicity* be given by

$$dm \equiv \frac{1}{N+1}(1+ \mid N^+ - N^- \mid) = \frac{1}{N+1}(1+ \mid F(\infty) - F(-\infty) \mid) = 1 - P(\pm\infty),$$
(17)

which is a random variable distributed over values $1 - \alpha_j$, $j = 0, 1, \ldots$; and $j \leq N/2$, as defined in (10). If the instrument z is monotonic, then $dm = 1$ a.s. and inference is based on $F(\beta) = F^+(\beta)$. If dm is not close to one, the instrument might be called *weak*. In particular, the weakness of an instrument is related to unbounded confidence sets $\mathcal{S}_j$. The set $\mathcal{S}_j$ is bounded if and only if $\pm\infty \notin \mathcal{S}_j$, which holds, using (12), if and only if $P(\pm\infty) < \alpha_j$. Consequently, using (17), $\mathcal{S}_j$ is bounded if and only if $1 - \alpha_j < dm$. So, dm is the smallest confidence level for which $\mathcal{S}_j$ is unbounded.

The weakness of an instrument is related to the 'first-stage regression' $x = z\pi + v$. Without making any assumptions in addition to the group invariance assumptions about $\varepsilon = y - x\beta$, we may define, analogous to $F(\beta)$, $\mathcal{S}_j$ and $P(\beta)$,

$$F^\pi(\pi) \equiv \# \left\{ i \mid \frac{z'(I_n - P_i)x}{z'(I_n - P_i)z} < \pi, i = 1, \ldots, N \right\},$$
(18)

$$\mathcal{S}_j^\pi \equiv \{\pi \mid F^\pi(\pi) \in \{j, \ldots, N - j\}\},$$
(19)

$$P^\pi(\pi) \equiv \max\{\alpha_j \mid \pi \in \mathcal{S}_j^\pi, j = 0, \ldots, N\} = \frac{N}{N+1}\left(1 - \left|1 - \frac{2F^\pi(\pi)}{N}\right|\right), \quad (20)$$

respectively. As $F^\pi(0) = N^-$, we find $P(\pm\infty) = P^\pi(0)$. We have proved the following:

PROPOSITION 3: *The following statements are equivalent:*

(a) $\mathcal{S}_j$ *is bounded,*   (b) $1 - \alpha_j < dm$,   (c) $0 \notin \mathcal{S}_j^\pi$.

To illustrate a low degree of monotonicity, consider $y = x\beta_o + \varepsilon$, where $\beta_o = 5$, and $x = z\pi_0 + v$, with $\pi_0 = 1$. A sample $n = 100$ was generated by $v_i = \varepsilon_i = (w_1 + w_2^2 + 3)_i^2$ and $z_i = \mid \log \mid \frac{w_3}{w_4} \mid\mid$, where for $i = 1, \ldots, 100$, $w \sim$ i.i.d.$\mathrm{N}(0, I_4)$. In order to make inferences about β_o, based on y, x and z, we only make a permutation invariance assumption about ε, conditional on z. Figure 2 gives the F-plot of β based on 1000 random permutations. The degree of monotonicity was 0.53, which seems too low to make this inference meaningful.

Figure 3 gives a histogram of all 1000 confidence limits, based on the same data. It shows a bimodal distribution, which is related to the bimodal distribution of the IV-estimator:[11]

$$\hat{\beta}_{IV} = \frac{E(z'(I_n - P)y \mid z, y)}{E(z'(I_n - P)x \mid z, x)} = \frac{z'(I_n - \iota_n(\iota_n'\iota_n)^{-1}\iota_n')y}{z'(I_n - \iota_n(\iota_n'\iota_n)^{-1}\iota_n')x} = \frac{z'(I_n - n^{-1}\iota_n\iota_n')y}{z'(I_n - n^{-1}\iota_n\iota_n')x},$$

where the expectation is over $P \in \mathcal{P}$. Contrary to the exact IV-estimator distribution, the distribution of the confidence limits can be computed as a function of the data.

---

[11]See Nelson and Startz (1990a, 1990b).

[Figure 2 about here.]

[Figure 3 about here.]

4.3 Bounded confidence sets and the power to reject underidentification

Dufour (1997) studies inference problems given "locally almost unidentified" parameters, e.g. parameters that are (nearly) not identified on certain subsets of the parameter space (see also references therein). These problems often arise in econometrics, and result in subsets of observationally equivalent parameter values, for instance when there are weak instruments. Dufour (1997) shows that the usual confidence sets for β must be unbounded with probability greater than or equal to $1 - \alpha$ when the coefficients on the exogenous variables in a standard structural model are rank-deficient. Confidence sets can also be empty. He interprets an empty confidence set as a rejection of the model itself, e.g. due to overidentifying restrictions. An unbounded confidence set for a structural coefficient suggests that the data may simply be uninformative about such coefficients, and (op. cit., p. 1383) "the occurrence of such a set may be interpreted as a symptom of the fact that the parameter cannot be precisely evaluated from the available data".

Moreover, Kleibergen (2002) analyzes the Angrist and Krueger (1991) dataset, and constructs 95% confidence sets for returns on education. For some specifications, he finds that the confidence set is *unbounded*, which indicates that instruments are weak, and that the amount of information contained therein is small (since the parameter can take any value, and so the confidence set contains all possible values). However, for other model specifications, Kleibergen finds *bounded* 95% confidence sets that seem sensible – and this suggests that there is some information

about returns to education in the instruments. Unbounded confidence sets may then be seen as helping to distinguish between weak and informative specifications.

A closely-related paper is Zivot et al. (1998, especially Section 4), and we thank a referee for drawing our attention to this. They link unboundedness of usual confidence sets with goodness-of-fit (e.g. F test) statistics for the first-stage regression. In particular (op. cit., p. 1130), they show that the Anderson-Rubin statistic is unbounded whenever the F test on the first-stage regression is insignificant. However, Zivot et al.'s (1998) treatment assumes normality of errors, as opposed to the weak assumptions made here.

If only "informative" (about β_o), i.e. bounded, confidence sets $\mathcal{S}_j$ are considered, their confidence level should be corrected for the probability of being bounded. That is, if the $(1 - \alpha)$-confidence set $\mathcal{S}_j$ is bounded with probability $p_\alpha = \mathrm{Prob}(\mathcal{S}_j$ is bounded), then a lower bound for the coverage probability, conditional on the boundedness of $\mathcal{S}_j$, is given by

$$
\begin{aligned}
\mathrm{Prob}(\beta_o \in \mathcal{S}_j \mid \mathcal{S}_j \text{ is bounded}) &= 1 - \mathrm{Prob}(\beta_o \notin \mathcal{S}_j \text{ and } \mathcal{S}_j \text{ is bounded})/p_\alpha \\
&\geq 1 - \alpha_j/p_\alpha.
\end{aligned}
\tag{21}
$$

Consequently, p_α is a relevant quantity for assessing the quality of inferences about β_o. The probability of the event in Proposition 3 equals p_α by definition. Clearly, making inferences about p_α amounts to making assumptions about x, which is the same as making assumptions about the variables in the 'first-stage regression', in addition to the group-invariance assumptions imposed on ε.

First, consider a simple assumption that states that x is a function of z. In that case, where we condition on z, N^- is not random and p_α equals either 1 or 0. So, there is no need to correct the confidence levels of sets $\mathcal{S}_j$ due to their

boundedness. A special case is given by linear regression, where $p_\alpha = 1$ if and only if $\alpha_j > 0$. Other assumptions describe x as a random vector conditional on z. This may affect the identifiability of β_o, and thus (cf. Dufour, 1997) it may affect inference on β_o. In particular, consider group invariance assumptions about the 'first-stage regression', $x = z\pi_0 + v$, such that the conditional distribution of v is not affected by linear transformations Pv, where $P \in \mathcal{P}$. This assumption excludes cases where the 'first-stage regression' is not necessarily linear (see e.g. Bekker, 1994).[12] Assumption-equivalent inference then amounts to $F^\pi(\pi_0) \sim U(\mathcal{U})$, where $F^\pi(\pi)$ is given in (18).

The first point to note is that β_o may not be identified. In other words, the group invariance assumption made about $y - x\beta_o$, conditional on z, may hold for *any* scalar β_o. In that case, the group invariance must also hold for the distribution of x. That is, x does not depend on z: $\pi = 0$. As a result, both $F^\pi(0)$ and dm, which equals $1 - P^\pi(0)$, will be uniformly distributed. Consequently, the probability of the event in Proposition 3 equals α_j. Hence $p_\alpha = \alpha_j$, and the lower bound (21) for the level of bounded confidence sets is not informative: $1 - \alpha_j/p_\alpha = 0$. Furthermore, a $(1 - \alpha_j)$-confidence set $\mathcal{S}_j$ will be bounded with probability α_j, $j = 0, 1, \ldots$; and $j \leq N/2$.

The null hypothesis of underidentification of β, i.e. $H_0 : \pi = \pi_0 = 0$, may be rejected with significance level α_j, if the $(1 - \alpha_j)$-confidence set $\mathcal{S}_j$ is bounded. Following rejection, inferences about β_o might be formulated based on bounded confidence sets, whose levels have been conditioned on their boundedness. If p_α is computed under the alternative hypothesis $H_a: \pi = \pi_a$, it equals the power of

---

[12]Note that the group-invariance assumptions about ε and v relate to their marginal distributions. A stronger assumption would be to assume such invariance with respect to $P(\varepsilon, v)$, $P \in \mathcal{P}$.

this test of underidentification.

Conditional on an equivalence class $\mathcal{C}_v = \{Pv \mid P \in \mathcal{P}\}$, and given π_a, this power can be computed by recognizing that, conditional on $\mathcal{C}_v$, x is distributed uniformly over the vectors $z\pi_a + P_i(x - z\pi_a)$, $i = 0, \ldots, N$. The choice $i = 0$ produces x, based on which N^- and dm have been computed. The choices $i = 1, \ldots, N$ produce alternatives for x, and hence for N^- and dm. Thus, the distribution of dm, conditional on $\mathcal{C}_v$, is generated. The power $p_{\alpha|\mathcal{C}_v}$, say, conditional on $\mathcal{C}_v$, is then found as the proportion of values dm that are larger than $1 - \alpha_j$. Of course, the true value π_0 would be unknown. Fortunately, we have a median unbiased estimator for π_0 given visually by the P-plot. Using this estimator, instead of the true value, provides an estimator for $p_{\alpha|\mathcal{C}_v}$ that is likely to have a very small median bias in relevant cases, where dm is close to one.

As a numerical example, artificial data were sampled in a similar manner to Kleibergen (2002). That is, $y = x\beta_o + \varepsilon$ and $x = z\pi_0 + v$, where $\beta_o = 0$ and $\pi = 0.1$. The disturbances ε_i and v_i have standard Gaussian distributions with a correlation equal to 0.99, and for $i = 1, \ldots, n$, the draws are independent; and $n = 500$. We assume permutation invariance of ε conditional on z. Figures 4 and 5 display, with restricted domain, the P-plots $P(\beta)$ and $P^\pi(\pi)$, based on 1000 random permutations. Here we find $dm = 0.98 = 1 - P^\pi(0)$ and the 95%-confidence set is indeed bounded, as indicated by the dotted line.

[Figure 4 about here.]

[Figure 5 about here.]

We also assume permutation invariance of v conditional on z. To verify that the probability of an unbounded 95%-confidence set is small, we computed

the quantity $p_{0.05|\mathcal{C}_{x-x\widehat{\pi}}} = \text{Prob}(dm > 0.95|\mathcal{C}_{x-z\widehat{\pi}}) = 0.6$, where $\widehat{\pi} = 0.11$ is the median unbiased estimator for π, i.e. the value for which $P^{\pi}(\pi)$ is maximal. So, the confidence level of the '95%-confidence set' needs to be corrected for the probability of being bounded. The corrected confidence level is approximately bounded by $1 - 0.05/0.60 = 0.92$.[13]

4.4 Nonlinear inferential functions

Instead of using a linear inferential function given by $g(\varepsilon) = z'\varepsilon$, which produces exact inference $F(\beta_o) \sim U(\mathcal{U})$, with $F(\beta)$ as given in (16), other functions could be considered as well. In particular, if the distribution of the disturbances has heavy tails it might be useful to choose, for example, $g(\varepsilon) = z'r_{\varepsilon}$, where r_{ε} contains the rank numbers of ε. In that case, when Condition 1 is satisfied, we still find $P(\pm\infty) = P^{\pi}(0)$ if $P^{\pi}(0)$ is computed based on a similar inferential function given by $g^{\pi}(v) = z'r_v$. If $\pi = 0$ can be rejected with more power based on this alternative inferential function, the confidence sets for β will be bounded with higher probability. Also note that in these cases the inference is based on a grid of values of β, although the reference set $\mathcal{R}(z'r_{y-x\beta})$ does not depend upon β.

If this approach is used in the numerical example of the previous section, there is no improvement due to the Gaussianity of the disturbances. However, if data are collected in a similar manner to the example used for Figure 2, there is a considerable improvement. As an example, consider inference on β_o, where $y = x\beta_o + \varepsilon$ and $x = z\pi_0 + v$. The data are based on $n = 1000$ draws from the same model that was used for Figure 2, where $n = 100$. Again, we assume permutation

---

[13]If exact probabilities p_{α} are computed based on full knowledge of the data generating process, we find $p_{0.05|\mathcal{C}_{x-z\pi_0}} = 0.51$ and $p_{0.05|z} = 0.62$.

invariance of ε conditional on z. Figures 6 and 7 give the P-plots $P(\beta)$ and $P^{\pi}(\pi)$, based on 1000 random permutations, both for the linear inferential function used earlier, and for the inferential function based on the rank numbers of ε. The improvement is striking.[14]

[Figure 6 about here.]

[Figure 7 about here.]

5   Assumption-equivalent inference and multiple instruments

This section considers inference based on multiple instruments. Assumption-equivalent inference is formulated in terms of joint confidence sets, which will be used to formulate conservative confidence intervals for separate elements of the parameter vector.

5.1   Anderson-Rubin-type confidence sets

First, we consider quadratic inferential functions $g(\varepsilon) = \varepsilon'Q\varepsilon$, where Q is a function of a matrix of instruments Z. To formulate invariance assumptions, we consider subgroups of the permutation-reflection group that satisfy Condition 1. Note that the reflection group contains too many elements, since $g(\varepsilon) = g(-\varepsilon)$. However, reflecting only $n-1$ elements may provide a suitable subgroup $\mathcal{P}$, for which all elements of the reference set $\mathcal{R}(g(\varepsilon))$ are different a.s.

Assumption-equivalent inference on β_o amounts to $F(\beta_o) \sim U(\mathcal{U})$, with $\varepsilon = f(y, X; \beta_o)$ and

---

[14]Here we find that $dm = 0.98$, and $\mathrm{Prob}(dm > 0.95 \mid \mathcal{C}_{x-z\hat{\pi}}) = 0.6$.

$$F(\beta) = \#\{P \mid \varepsilon'Q\varepsilon < \varepsilon'P'QP\varepsilon, P \in \mathcal{P}\}. \tag{22}$$

In particular, we consider matrices Q such that relevant confidence sets are given by $\mathcal{H}_j = \mathcal{C}(\mathcal{V})$, as in (3), where $\mathcal{V} = \{j, \ldots, N\}$.

In the context of the linear model, where $\varepsilon = y - X\beta_o$, an intuitively appealing choice for Q would be $Q = P_Z = Z(Z'Z)^{-1}Z$. We assume the matrix $(y, X)'P'(I_n - P_Z)P(y, X)$ has full rank for $P \in \mathcal{P}$ a.s. Then, the same function $F(\beta)$ is found a.s. for the inferential function given by

$$g(\varepsilon) = \frac{\varepsilon'P_Z\varepsilon}{\varepsilon'(I_n - P_Z)\varepsilon}. \tag{23}$$

Alternatively, in the presence of heavy tails, one might consider using the rank numbers r_ε instead of ε.

If the matrix $(y, X)'P_Z(y, X)$ has full rank a.s., then $g(y - X\beta)$ has a positive minimum and a finite maximum a.s. As a result, elementary confidence sets $\mathcal{C}(\{i\})$ may be empty for small and large values of $i = 0, 1, \ldots, N$. Consequently, confidence sets $\mathcal{H}_j$ may be empty or unbounded for values $j > 0$. For the just-identified case, where the matrix $(y, X)'P_Z(y, X)$ has a one-dimensional null space, the minimum of $g(y - X\beta)$ equals zero. In that case, confidence sets $\mathcal{H}_j$ can be unbounded, but they cannot be empty. Furthermore, for finite groups, the construction of confidence sets is a matter of trial and error: $\beta \in \mathcal{H}_j$ if $F(\beta) \in \{j, \ldots, N\}$.

The computation of confidence sets $\mathcal{H}_j$ is easier for the infinite rotation group than it is for the finite permutation-reflection group. That is, for the rotation

group, the reference set $\mathcal{R}(g(y - X\beta))$ does not depend on y, X or β.[15] When multiplied by $(n- \text{tr}(P_Z))/\text{tr}(P_Z)$, the reference set has an F-distribution with $\text{tr}(P_Z)$ and $n- \text{tr}(P_Z)$ degrees of freedom.

If β_o refers only to the coefficients of endogenous variables, i.e. $\varepsilon = y - X\beta_o - Z_1\gamma$, then inference on β_o can also be based on

$$g(\varepsilon) = \frac{\varepsilon'P_{Z^*}\varepsilon}{\varepsilon'(I_n - P_Z)\varepsilon} = \frac{\varepsilon'P_{Z^*}\varepsilon}{\varepsilon'(I_n - P_{Z_1} - P_{Z^*})\varepsilon}, \tag{24}$$

where $Z^* = (I_n - P_{Z_1})Z_2$ and $Z = (Z_1, Z_2)$. Note that $g(\varepsilon) = g(y - X\beta_o)$, and that the reference set $\mathcal{R}(g(y - X\beta - Z_1\gamma))$ does not depend on y, β or γ. Inference of this form, based on the rotation group, amounts to Anderson-Rubin confidence sets (cf. Anderson and Rubin, 1949; Bartlett, 1948), which are exact under normality. However, contrary to the classical regression context, where $Z_2 = X$, the Anderson-Rubin confidence sets need neither be bounded nor convex, and they might be empty in case of overidentification.[16]

---

[15]This property is shared by the finite groups when rank numbers r_ε are used instead of ε.

[16]This paper can be seen as complementary to recent work on similar tests in instrumental variables regression. For instance, Kleibergen's (2002) K-statistic is an alternative to the Anderson-Rubin (AR) test. It is asymptotically pivotal, and converges uniformly in distribution to a χ^2 with d.g.f. equal to the number of explanatory variables, and shows improved power properties over AR in the overidentified case. An inferential function analogous to the K-statistic depends on both ε and X. Instead of (23), we have $g(\varepsilon, X) = \varepsilon'P_{\tilde{Z}}\varepsilon / \varepsilon'(I_n - P_Z)\varepsilon$, with $\tilde{Z} = P_Z \{X - \varepsilon\varepsilon'(I_n - P_Z)X / \varepsilon'(I_n - P_Z)\varepsilon\}$. Since X is not independent of ε, it is not possible to describe the distribution of $g(\varepsilon, X)$ conditional on an equivalence class $\mathcal{C}_\varepsilon$. If $\Pi = E(X|Z)$ were known then an equivalence class $\{P(\varepsilon, V) \mid P \in \mathcal{P}\}$ could be used, where $V = X - Z\Pi$. In fact, Bekker and Kleibergen (2003) describe bounds for the exact distribution of the K-statistic under Gaussianity where one of the bounds is found for $\Pi = 0$. Here, exact inference based on group-invariance assumptions, such as the exchangeability of the rows of (ε, V), can be formulated for the extreme case where $\Pi = 0$. An alternative similar procedure was developed by Moreira (2003), and is exact under normality, with known covariance matrix. When the error distribution is unknown, Moreira shows that modified versions of similar tests based on conditional distributions are asymptotically similar under Staiger and Stock's (1997) weak instrument asymptotics, and approximately similar in moderate-sized samples. A conditional pseudo-likelihood ratio test is shown to have good power properties under weak identification. However, the K-statistic and the conditional LR statistic have limiting χ^2 and nonstandard distributions respectively, as opposed to the exact symmetry-based procedures developed here.

5.2   Group-invariance assumptions in a dynamic context

An example of exact inference in a dynamic context is discussed by Dufour and Kiviet (1998). Consider the first-order autoregressive distributed lag model given by $y_t = \beta_o y_{t-1} + x_t' \gamma + \varepsilon_t$, $t = 1, \ldots, n$, where $\varepsilon \sim N(0, \sigma^2 I_n)$, and x_t contains strongly exogenous regressors. Let the regression equation in vector notation be given by

$$y = y_{-1}\beta_o + Z_1\gamma + \varepsilon. \tag{25}$$

The 'first-stage' regression equation follows by recurrent substitution as

$$y_{-1} = y_0\kappa(\beta_o) + C(\beta_o)Z_1\gamma + C(\beta_o)\varepsilon,$$

where

$$\kappa(\beta) \;=\; \begin{pmatrix} 1 \\ \beta \\ \beta^2 \\ \vdots \\ \beta^{n-1} \end{pmatrix}, \; C(\beta) = \begin{pmatrix} 0 & \cdot & \cdot & \cdot & 0 \\ 1 & 0 & & & \cdot \\ \beta & 1 & 0 & & \cdot \\ \vdots & & \ddots & & \cdot \\ \beta^{n-1} & \ldots & \beta & 1 & 0 \end{pmatrix}.$$

Dufour and Kiviet (1998) assume that the distribution of ε is known up to a scale factor. They base their inference procedures on the more general model

$$y = y_{-1}\beta_o + Z\lambda + \varepsilon; \tag{26}$$

$$Z = (Z_1, Z_2),$$

$$Z_2 = (\kappa(\beta_o), C(\beta_o)Z_1),$$

where $\lambda = (\gamma', \lambda_2, \lambda_3')'$. Equation (26) reduces to (25) when $\lambda_2 = 0$ and $\lambda_3 = 0$. A monotonic transformation of the likelihood-ratio statistic, under normality, for testing $H_o : \beta = \beta_o$, $\lambda_2 = 0$ and $\lambda_3 = 0$ is given by (cf. Dufour and Kiviet, 1998, (2.25))

$$g(\varepsilon) = \frac{\varepsilon' P_{\tilde{Z}^*} \varepsilon}{\varepsilon'(I_n - P_{\tilde{Z}})\varepsilon} = \frac{\varepsilon' P_{\tilde{Z}^*} \varepsilon}{\varepsilon'(I_n - P_{Z_1} - P_{\tilde{Z}^*})\varepsilon} = \frac{\varepsilon'(I_n - P_{Z_1})\varepsilon}{\varepsilon'(I_n - P_{\tilde{Z}})\varepsilon} - 1; \quad (27)$$

$$\tilde{Z} = (Z_1, \tilde{Z}_2),$$

$$\tilde{Z}_2 = (Z_2, C(\beta)\varepsilon),$$

$$\tilde{Z}^* = (I_n - P_{Z_1})\tilde{Z}_2.$$

We see that both $\tilde{Z}$ and $\tilde{Z}^*$ depend on ε, which is not problematic for generating the distribution of $g(\varepsilon)$ conditional on equivalence classes – this observation follows as a consequence of the results of Dufour and Kiviet (1998). Also, $g(\varepsilon) = g(y - y_{-1}\beta_o)$ and so the reference set $\mathcal{R}(g(\varepsilon))$ does not depend on ε if $\mathcal{P}$ is the rotation group. Thus, a rotation invariance assumption about ε, conditional on Z_1, implies exact inference $F(\beta_o) \sim U([0,1])$, where $F(\beta)$ is based on the inferential function given in (27).

Nonparametric generalizations can now easily be formulated based on group invariance assumptions related to the finite permutation-reflection group. For first-order autoregressive distributed lag models, inference can be based on (27). However, for these nonparametric generalizations the reference sets do generally depend both on the data and on the full vector of regression coefficients. Alternatively,

other inferential functions could be considered where some, or all vectors ε in (27) are replaced by rank numbers r_ε, or some other function.

We might also consider formulating nonparametric estimators for β_o. That is, if $\mathcal{P}$ is the rotation group, the value β that maximizes $F(\beta)$ in (22), where $Q = P_Z$, is the limited information maximum likelihood (LIML) estimator. Similar to the Hodges-Lehmann estimator, it is the value most supported by the confidence sets $\mathcal{R}_j$, $j = 0, 1, \ldots, N$. As a nonparametric generalization of LIML, we might consider an element of the set that maximizes (22) if $\mathcal{P}$ is not the rotation group. This approach can be applied to other cases as well, including cases where rank numbers are used instead of ε.

5.3 An application to nonlinear regression models

Inference of the form (22) is not restricted to linear models. For a nonlinear model we could use $Q = P_Z$, where

$$Z = \partial f(y, X; \beta)/\partial \beta'. \tag{28}$$

As an example, consider the simple nonlinear regression model

$$y_i = \beta_{o1} x_i^{\beta_{o2}} + \varepsilon_i, \;\; i = 1, \ldots, n. \tag{29}$$

Based on group invariance assumptions, exact two-dimensional confidence sets follow from $F(\beta_o) \sim U(\mathcal{U})$, where $F(\beta)$ and Z are given in (22) and (28), respectively. In particular, if the disturbances ε_i are independent and symmetrically distributed conditional on x, we may consider the reflection subgroup with elements P_i, $i = 0, \ldots, 2^{n-1} - 1$, for which the first diagonal element equals

one: $P_{i11} = 1$.

Consider data where the elements x_i are independent and $x_i \sim \chi_1^2$; the elements ε_i are independent and $\varepsilon_i = w_i^3$, with $w_i \sim (U([0, |x_i^*|]) - 0.5|x_i^*|)$, $i = 1, \ldots, 100$. So, there is both heteroscedasticity and non-Gaussianity. Using an reflection invariance assumption about ε, conditional on $x_1, \ldots, x_{100}$, and based on a 501×501 grid and 2500 random reflections, Figure 8 gives exact joint confidence sets for $\beta_{o1} = 0.5$ and $\beta_{o2} = 1$.

[Figure 8 about here.]

Based on such joint confidence sets, conservative confidence intervals for a separate element β_{o2}, say, can be bounded by the minimum and maximum of β_2, over a sufficiently large grid, such that $F(\beta) \in \{j, \ldots, N\}$. However, it might be useful to consider different inferential functions $g(\varepsilon)$ for inferences on different parameters.

Analogous to t-inference, which is based on rotational invariance, an inferential function of the form

$$g(\varepsilon) = \frac{z_2'(I_n - P_{z_1})\varepsilon}{(\varepsilon'(I_n - P_Z)\varepsilon)^{1/2}}, \tag{30}$$

might be useful to formulate inference on β_{o2}. When comparing conservative confidence intervals for β_{o2}, based on joint confidence sets generated by (23) and (30), respectively, the latter may perform better.

Given the data used for Figure 8, inference based on (30) and reflection invariance is displayed in Figure 9. Indeed, if β_{o1} and β_{o2} are known to be restricted to [0,1] and [0.6,1.6], respectively, a comparison of conservative 95% confidence sets for β_{o2} in Figure 8, [0.786,1.46], and Figure 9, [0.850, 1.422], favours the latter.

[Figure 9 about here.]

6   Conclusion

The paper provides exact inference on the parameters of the functional form for a variety of econometric models. The inference is based on mild symmetry assumptions about the distribution of disturbances. Exact inference based on normality can be derived in a similar way. When applied to the linear model, the method provides a unified approach to well-known parametric and nonparametric tests. Especially when asymptotic methods break down, this new approach may be useful. For example, when sample sizes are small, when distributions of disturbances or explanatory variables are heavy-tailed, or when parameter points are close to underidentification, or when dynamic models are explosive, the proposed methods remain exact.

Some special attention has been given to the relation between bounded confidence sets and the power to reject a hypothesis of underidentification. When parameter points can be underidentified, exact confidence sets will be unbounded with positive probability. Therefore, if the hypothesis of underidentification is rejected and confidence sets are bounded, the coverage probability of confidence sets must be corrected for the probability of being bounded. Our approach provides an estimate of such a correction of the nominal size.

This leaves open the question of optimality. Usually optimality is formulated in asymptotic terms. Not infrequently this leads to 'optimal' inference that performs poorly in small samples. Whether or not it is possible to formulate general guidelines for formulating inferential functions with desirable properties is a matter for further research. We used intuitively motivated linear and

quadratic functions, as well as functions based on test statistics derived under normality. This does not exhaust all possibilities. In an example with heavy-tailed distributed disturbances, inference based on rank numbers of the disturbances was shown to greatly improve the quality of the inference when compared to a linear function of the disturbances.

A   Appendix

If $F(\beta_o) \sim U(\mathcal{U})$, where $\mathcal{U} = \{0, 1, \ldots, N\}$, then exact inference can also be based on a random sample of M, say, matrices $P \in \mathcal{P}$. Let $F_M(\beta)$ indicate the function F when it is computed based on these M matrices instead of the full group $\mathcal{P}$. It will be shown that $F_M(\beta_o) \sim U(\{0, 1, \ldots, M\})$ if the sample is taken without replacement. If the sample is taken with replacement, then $\mathrm{Prob}(F_M(\beta_o) = j) = r_j$, where

$$r_j = \frac{1}{(N+1)} \sum_{k=0}^{N} \left(\begin{array}{c} M \\ j \end{array} \right) (k/N)^j (1 - k/N)^{M-j}, \ j = 0, \ldots, M. \qquad \text{(A1)}$$

Let a_i, $i = 1, \ldots, N$, be random variables equal to either 1 or 0, and let the probability distribution of $\sum_{i=1}^{N} a_i$ over $\{0, \ldots, N\}$ be given by $(p_0, \ldots, p_N)' = p$, say. Consider a random drawing a from $\{a_1, \ldots, a_N\}$, and let the probability distribution of $\sum_{i=1}^{N} a_i - a$ over $\{0, \ldots, N-1\}$ be given by $(q_0, \ldots, q_{N-1})' = q$, say. Then, for $j = 0, \ldots, N-1$,

$$q_j \;=\; \mathrm{Prob}\left(a = 0 \mid \sum_{i=1}^{N} a_i = j \right) p_j + \mathrm{Prob}\left(a = 1 \mid \sum_{i=1}^{N} a_i = j + 1 \right) p_{j+1}$$

$$= \; (1 - j/N)p_j + ((j+1)/N)p_{j+1}.$$

Let

$$
S_N \;=\; N^{-1}
\begin{bmatrix}
N & 1 & & & & \\
 & N-1 & 2 & & \emptyset & \\
 & & \ddots & \ddots & & \\
 & \emptyset & & 2 & N-1 & \\
 & & & & 1 & N
\end{bmatrix},
$$

then $q = S_N p$, which is the probability distribution of the sum of a random sample of size $N-1$ taken without replacement from $\{a_1, \ldots, a_N\}$. Similarly, for a random sample of size M, we find a probability distribution given by $S_{M+1} S_{M+2} \ldots S_N p$. Clearly this latter distribution is uniform if the distribution given by p is uniform. In particular, let $a_i = 1$ if $g(\varepsilon) < g(P_i \varepsilon)$, and $a_i = 0$ otherwise, $i = 1, \ldots, N$. Then $S_{M+1} S_{M+2} \ldots S_N p$ describes the uniform distribution of $F_M(\beta_o)$ over $\{0, 1, \ldots, M\}$ when the sample is taken without replacement.

Let the sum s_M, say, of a random sample of size M taken with replacement from $\{a_1, \ldots, a_N\}$ have a probability distribution over $\{0, \ldots, M\}$ equal to $(r_0, \ldots, r_M)$. Then, for $j = 1, \ldots, M$,

$$
r_j \;=\; \sum_{k=0}^{N} \mathrm{Prob}\left(s_m = j \mid \sum_{i=1}^{N} a_i = k \right) p_k = \sum_{k=0}^{N} \binom{M}{j} (k/N)^j (1 - k/N)^{M-j} p_k.
$$

Consequently, if $p_k = 1/(N+1)$, we find (A1).

Aigner, D. J., Hsiao, C., Kapteyn, A. and T. J. Wansbeek, 1984, Latent variable models in econometrics, in: Z. Griliches and M. D. Intriligator, eds., Handbook of econometrics, Vol. 2 (North-Holland, Amsterdam) 1321-1393.

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., 1993, Exactly median-unbiased estimation of first-order autoregressive/unit root models, Econometrica 61, 139-165.

Andrews, D. W. K. and P. C. B. Phillips, 1987, Best median-unbiased estimation in linear regression with bounded asymmetric loss functions, Journal of the American Statistical Association 82, 886-893.

Angrist, J. D. and A. B. Krueger, 1991, Does compulsory school attendance affect schooling and earnings?, Quarterly Journal of Economics 106, 979-1014.

Bartlett, M. S., 1948, A note on the statistical estimation of demand and supply relations from time series, Econometrica 16, 323-329.

Basu, D., 1980, Randomization analysis of experimental data: the Fisher randomization test, Journal of the American Statistical Association 75, 575-595 (with discussion).

Bekker, P. A., 1994, Alternative approximations to the distributions of instrumental

variable estimators, Econometrica 62, 657-681.

Bekker, P. A., 2002, Exact inference for the linear model with groupwise heteroscedastic spherical disturbances, Journal of Econometrics 111, 285-302.

Bekker, P. A. and F. Kleibergen, 2003, Finite-sample instrumental variables inference using an asymptotically pivotal statistic, Econometric Theory 19, 744-753.

Bound, J., Jaeger, D. A. and R. M. Baker, 1995, Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak, Journal of the American Statistical Association 90, 443-450.

Dawid, A. P., 1988, Symmetry models and hypotheses for structured data layouts, Journal of the Royal Statistical Society B 50, 1-34 (with discussion).

Donald, S. G. and W. K. Newey, 2001, Choosing the number of instruments, Econometrica 69, 1161-1191.

Dufour, J. -M., 1997, Some impossibility theorems in econometrics with applications to structural and dynamic models, Econometrica 65, 1365-1387.

Dufour, J. -M. and J. F. Kiviet, 1998, Exact inference methods for first-order autoregressive distributed lag models, Econometrica 66, 79-104.

Efron, B., 1969, Student's t-test under symmetry conditions, Journal of the American Statistical Association 64, 1278-1302.

Gabriel, K. R. and W. J. Hall, 1983, Rerandomization inference on regression and shift effects: computationally feasible methods, Journal of the American Statistical Association 78, 827-836.

Hahn, J. and J. Hausman, 2002, A new specification test for the validity of instrumental variables, Econometrica 70, 163-189.

Hahn, J. and J. Hausman, 2003, Weak instruments: diagnosis and cures in empirical econometrics, American Economic Review Papers and Proceedings 93, 118-125.

Hájek, J., 1969, Nonparametric statistics (Holden-Day, San Francisco).

Hodges, J. L., Jr. and E. L. Lehmann, 1963, Estimates of location based on rank tests, Annals of Mathematical Statistics 34, 598-611.

Imbens, G. and P. Rosenbaum, 2005, Robust, accurate confidence intervals with a weak instrument: quarter of birth and education, Journal of the Royal Statistical Society A 168, 109-126.

Kendall, M. G., 1938, A new measure of rank correlation, Biometrika 30, 81-93.

Kendall, M. G., 1975, Rank correlation methods, 4th Edition (Griffin, London).

Kleibergen, F., 2002, Pivotal statistics for testing structural parameters in instrumental variables regression, Econometrica 70, 1781-1803.

Kleibergen, F., 2004, Testing subsets of structural parameters in the instrumental variables regression model, Review of Economics and Statistics 86, 418-423.

Kroese, A. H. and W. Schaafsma, 1998, Distributional inference, in: Kotz, S., Read, C. B. and N. L. Johnson, eds., Encyclopedia of statistical sciences, Vol. 2 (John Wiley and Sons, New York) 188-192.

Lehmann, E. L., 1975, Nonparametrics: statistical methods based on ranks (Holden-Day, San Francisco).

Lehmann, E. L., 1997, Testing statistical hypotheses, 2nd Edition (Springer-Verlag, New York).

Mann, H. B., 1945, Nonparametric tests against trend, Econometrica 13, 245-259.

Maritz, J. S., 1995, Distribution-free statistical methods, 2nd Edition (Chapman and Hall, London).

Moreira, M. J., 2001, Tests with correct size when instruments can be arbitrarily weak, unpublished manuscript, Harvard University.

Moreira, M. J., 2003, A conditional likelihood ratio test for structural models, Econometrica 71, 1027-1048.

Nelson, C. R. and R. Startz, 1990a, The distribution of the instrumental variables estimator and its t-ratio when the instrument is a poor one, Journal of Business 63, S125-S140.

Nelson, C. R. and R. Startz, 1990b, Some further results on the exact small sample properties of the instrumental variable estimator, Econometrica 58, 967-976.

Robinson, J., 1987, Nonparametric confidence intervals in regression: the bootstrap and randomization methods, in: Puri, M. L., Vilaplana, J. P. and W. Wertz, eds., New perspectives in theoretical and applied statistics (Wiley, New York) 243-255.

Staiger, D. and J. H. Stock, 1997, Instrumental variables regression with weak instruments, Econometrica 65, 557-586.

Stock, J. H., Wright, J. H. and M. Yogo, 2002, A survey of weak instruments and weak identification in generalized method of moments, Journal of Business and Economic Statistics 20, 518-529.

Theil, H., 1950, A rank-invariant method of linear and polynomial regression analysis I, Proceedings of Koninklijke Nederlandse Akademie van Wetenschappen A 53, 386-392.

Tritchler, D., 1984, On inverting permutation tests, Journal of the American Statistical Association 79, 200-207.

Wald, A., 1940, The fitting of straight lines if both variables are subject to error, Annals of Mathematical Statistics 11, 284-300.

Wang, J. and E. Zivot, 1998, Inference on structural parameters in instrumental variables regression with weak instruments, Econometrica 66, 1389-1404.

Woglom, G., 2001, More results on the exact small sample properties of the instrumental variable estimator, Econometrica 69, 1381-1389.

Zivot, E., Startz, R. and C. R. Nelson, 1998, Valid confidence intervals and inference in the presence of weak instruments, International Economic Review 39, 1119-1146.

Figure 1: The P-plot for $\beta_o = 1$, $n = 10$ ($P(\beta)$ against β).

Figure 2: F-plot for $\beta_o = 5$, $n = 100$; $dm = 0.53$ ($F(\beta)/N$ against β).

Figure 3: Histogram of confidence limits: $\beta_o = 5$, $n = 100$; $dm = 0.53$ (frequency of confidence limits against β).

Figure 4: P-plot for $\beta_o = 0, \pi_o = 0.1; n = 500$ ($P(\beta)$ against β).

Figure 5: P-plot for $\pi_o = 0.1; n = 500$ ($P^\pi(\pi)$ against π).

Figure 6: P-plot for $\beta_o = 5, \pi_0 = 1; n = 1000$ ($P(\beta)$ against β). Solid line: inference based on ranks. Dashed line: linear inferential function.

Figure 7: P-plot for $\pi_0 = 1; n = 1000$ ($P^\pi(\pi)$ against π). Solid line: inference based on ranks. Dashed line: linear inferential function.

Figure 8: Joint confidence sets for $\beta_{o1} = 0.5$, $\beta_{o2} = 1$; $n = 100$; sizes: 0.99, 0.95, 0.5, 0.05 (β_2 against β_1).

Figure 9: Joint confidence sets for $\beta_{o1} = 0.5$, $\beta_{o2} = 1$; $n = 100$; sizes: 0.99, 0.95, 0.5, 0.05 (β_2 against β_1).
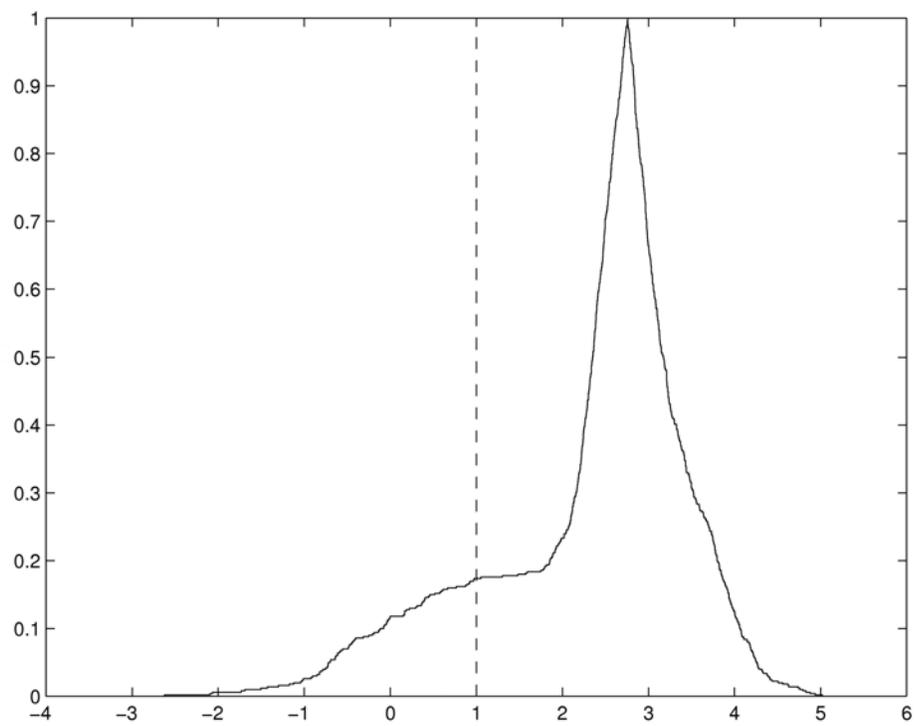
Figure 1

Figure 2

Figure 3

Figure 4

P (+/- infinity)

Figure 5

Figure 6

P (+/- infinity)
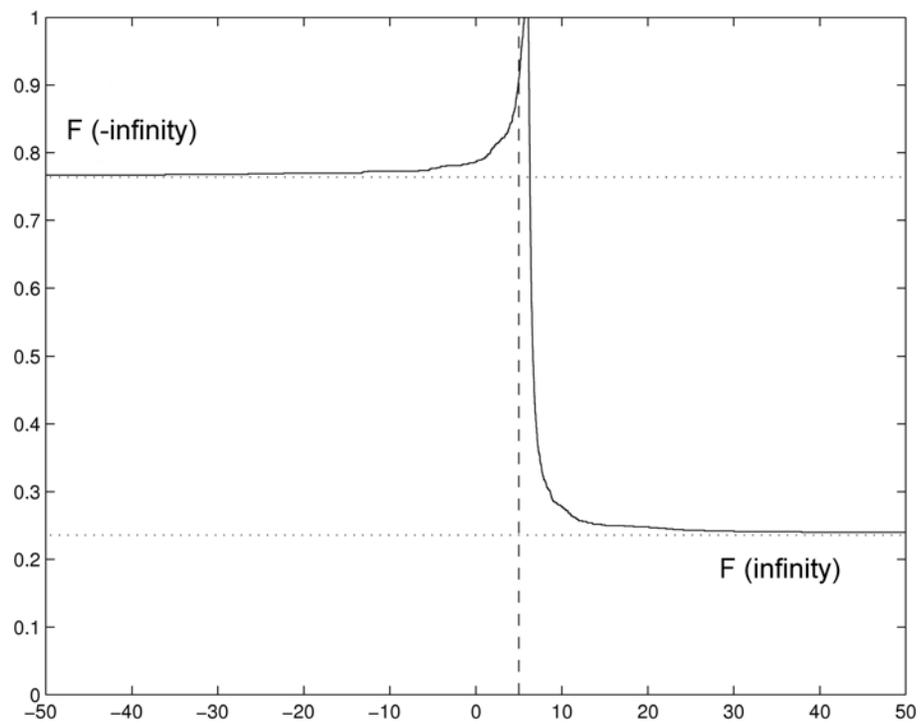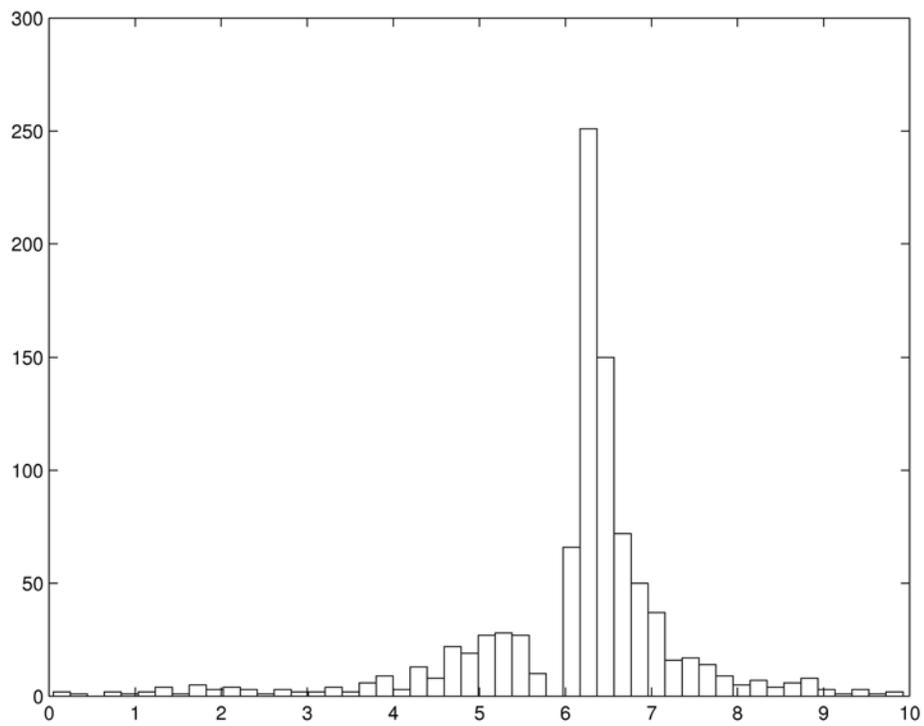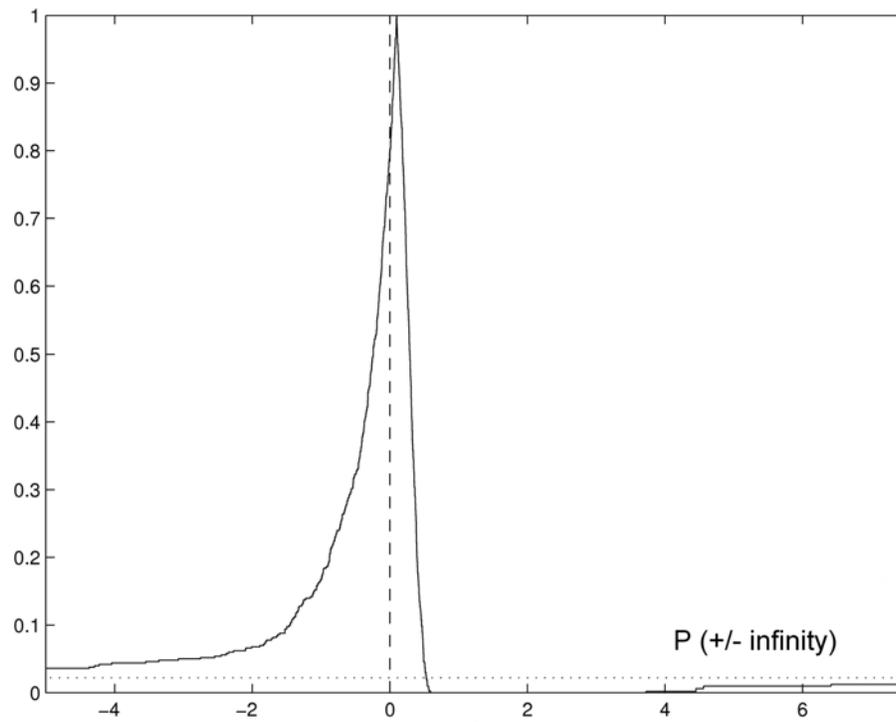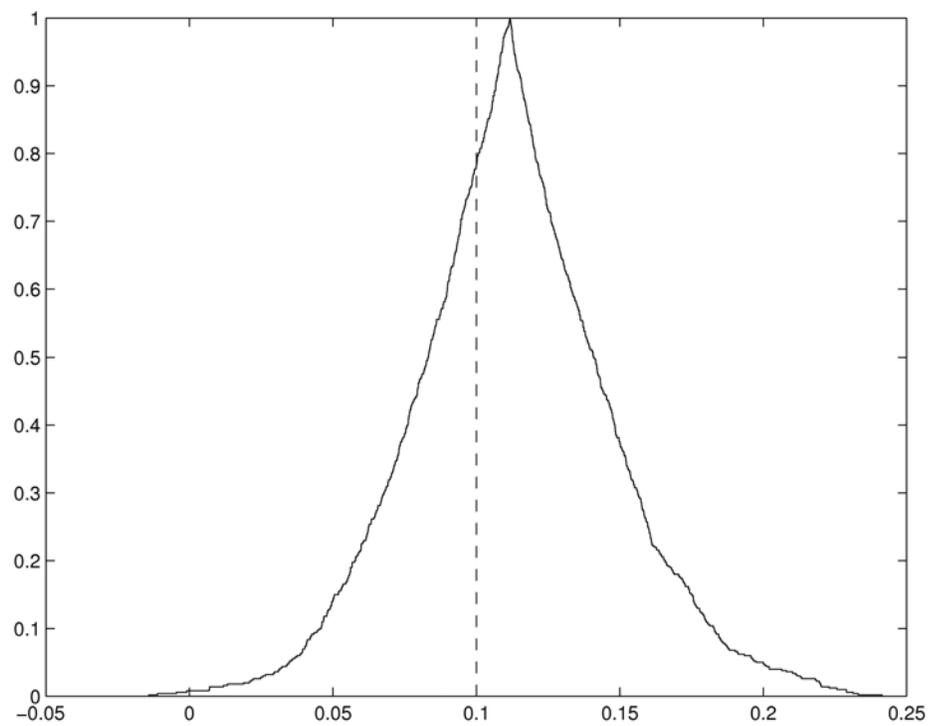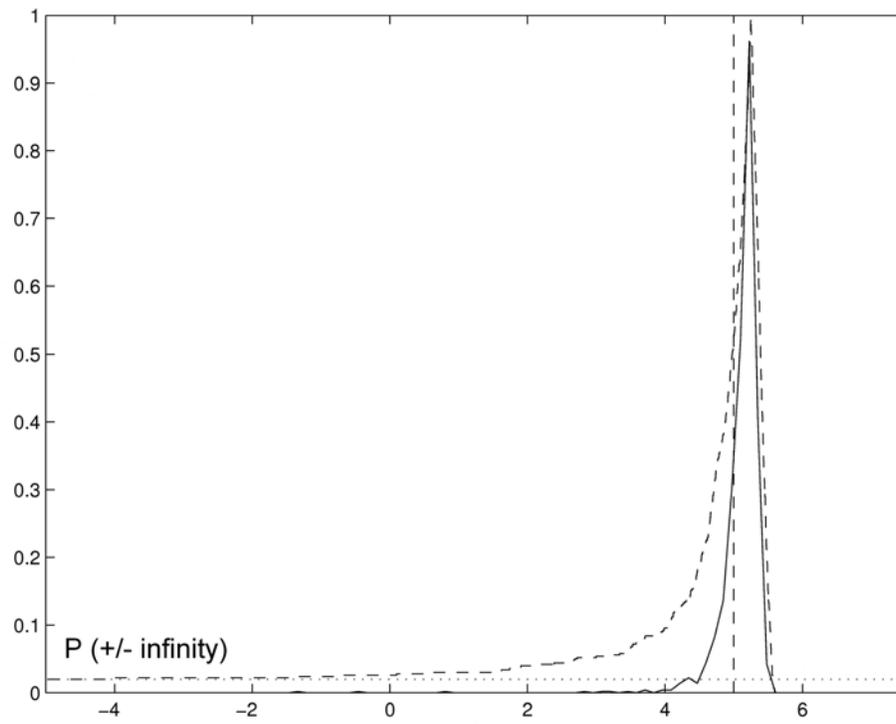
Figure 7

Figure 8

Figure 9