INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2010-56

Finite and Large Sample Distribution-Free Inference in Median Regressions with Instrumental Variables

E. COUDIN[1] – J.-M. DUFOUR[2]

---

1. CREST-ENSAE. *Mailing address* : CREST, Laboratoire de Microéconométrie, 15 Boulevard Gabriel Péri, 92245 Malakoff, Cedex, France. Email : elise.coudin@ensae.fr

2. William Dow Professor of Economics, McGill University, Montréal, Canada.
Email : jean-marie.dufour@mcgill.ca Web page : http://www.jeanmariedufour.com

Finite and large sample distribution-free inference in median regressions with instrumental variables *

Elise Coudin † Jean-Marie Dufour ‡

ENSAE-CREST McGill University

First version: April 2005

Revised: February 2008

This version: December 2010

* The authors thank Marine Carrasco, Frédéric Jouneau, Marc Hallin, Thierry Magnac, Bill McCausland, Benoît Perron, Alain Trognon for useful comments and constructive discussions. This paper was presented at the 2008 Econometric Society European Meeting (Milano, August 2008). A first version of this paper was written when the authors were at the Université de Montréal. This work was supported by the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Alexander-von-Humboldt Foundation (Germany), the Institut de Finance mathématique de Montréal (IFM2), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Canada Council for the Arts (Killam Fellowship), the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, the Fonds de recherche sur la société et la culture (Québec), and the Fonds de recherche sur la nature et les technologies (Québec).

† Centre de recherche en économie et statistique (CREST-ENSAE). Mailing address: Laboratoire de microéconometrie, CREST, Timbre J390, 15 Bd G. Péri, 92245 Malakoff Cedex, France. E-mail: elise.coudin@ensae.fr.

‡ William Dow Professor of Economics, McGill University. Centre interuniversitaire de recherche en analyse des organisations (CIRANO), Centre interuniversitaire de recherche en économie quantitative (CIREQ), and Department of Economics, McGill University. Mailing address: Department of Economics, McGill University, 855 Sherbrooke Street West, Montreal, Quebec H3A 2T7, Canada . TEL: 1 514 398 8879; FAX: 1 514 398 4850; e-mail: jean-marie.dufour@mcgill.ca. Web page: http://www.jeanmariedufour.com.

ABSTRACT

This paper develops a finite-sample distribution-free inference system for the parameters of a structural nonlinear model. We introduce an instrument validity condition with respect to the structural error signs. We notice that the conditional distribution of the structural error signs is nuisance-parameter free and known. This allows us to conduct a Monte-Carlo-based inference which, in conjunction with projection techniques, produces valid results in finite samples robust to identification failures in very general settings - nonnormality, heteroskedasticity, nonlinearly dependent errors, weak instruments. The proposed inference method is asymptotically valid in presence of serially dependent errors. Basically, the sign-based approach relies on artificial regressions where the signs of the constrained residuals are regressed on some "auxiliary" instruments [Anderson and Rubin (1949), Dufour (2003)]. Then, we study the problem of building optimal instruments, in case of overidentification. We provided IV sign-based estimators in identifed setups. Consistency and asymptotical normality are established under weaker assumptions than the ones used for the 2SLAD estimator asymptotic theory. Finally, simulations show that sign-based methods overcome usual methods and methods robust to weak instruments in non-normal and heteroskedastic settings. A re-analysis of the returns to education based on Angrist and Krueger (1991) data is also provided.

Key words: sign-based methods; median regressions; instrumental variables; finite samples; simultaneous inference; Monte Carlo tests; projection methods; non-normality; heteroskedasticity; serial dependence; GARCH; stochastic volatility; bootstrap.

Contents

List of Propositions and Theorems

List of Tables

1. Introduction

Instrumental Variable (IV) regression results greatly rely on the quality of the instruments used. When the latter are weakly correlated with the endogenous variable, usual estimators are biased and asymptotic approximations are not anymore valid; see Bound, Jaeger, and Baker (1995), Staiger and Stock (1997), Dufour (1997, 2003), Wang and Zivot (1998), Stock and Wright (2000). Inference relying on estimator asymptotic behavior such as Wald tests may be greatly misleading. One approach to circumvent the problem of weak instruments is to dissociate testing from estimation and to investigate alternative test procedures. Contrary to Wald tests, tests based on the Anderson-Rubin (AR) statistic have correct size for normally distributed disturbances without requiring the parameter to be identified. AR tests are valid in the presence of weak instruments; see Anderson and Rubin (1949), Dufour (1997), Nelson, Startz, and Zivot (1998). However, the AR procedure relies on a Gaussian assumption or at least on some asymptotic justification. In small samples with non-Gaussian disturbances, AR tests (such as any asymptotic test) may be affected by size distortions. Fully exact inference procedures in models where some regressors are endogenous have been less studied. In a regression setup, we propose to use the residual signs to conduct nonparametric valid tests with controlled level for any sample size.

We consider here a possibly nonlinear equation which involves endogenous regressors. A set of exogenous variables is available and no parametric assumption is imposed on the disturbance process. The latter is only assumed to have median zero conditional on the exogenous variables (hereafter, the instruments) and its own past. Without any further restriction, we notice that the sign vector distribution of the constrained residuals is a pivotal function. This property is actually a natural extension of the one stated in Coudin and Dufour (2009). The sign vector distribution does not depend on nuisance parameters and can easily be simulated. Basically, we use Monte Carlo test techniques [see Dwass (1957), Barnard (1963) and Dufour (2006)] to construct joint sign-based tests that control the level for any sample size. The validity of these tests does not depend on identification assumptions nor on any parametric approximation. In the presence of weak instruments or identification failures, sign-based test levels still equal their nominal size. Then, a complete system of finite-sample inference - as well as asymptotic extensions - can be applied [see Coudin and Dufour (2009)]. Simultaneous confidence sets for the whole parameter are obtained by test inversion. Next, confidence sets and tests of general hypotheses are built using projection techniques [see Dufour and Kiviet (1998), Dufour and Jasiak (2001), Dufour and Taamouti (2005)]. Finally, Hodges-Lehmann estimators are provided in identified cases [Hodges and Lehmann (1963), Coudin and Dufour (2008)]. They correspond to the parameter value least rejected by the tests. As entailed by the results in Dufour (1997), the derived confidence regions may have a non-zero probability of being unbounded in the presence of identification failures.

Nonparametric approaches investigated up to now in the literature have been based on rank and permutation tests. A rank-version of the AR test was introduced by Andrews and Marmer (2008). It dominates the usual AR in terms of size and power for asymmetric and thick tail error distributions. It yields exact tests if the exogenous regressors are independent of instruments and errors. Besides, Bekker and Lawford (2008) proposed exact inference based on permutation tests. Both methods are especially adapted to cross-sectional data, since the errors are assumed to be independent and identically distributed ($i.i.d.$). By contrast, sign-based methods are known to be the only way of producing inference procedures that are proved to be valid under heteroskedasticity of unknown form for a given sample size; see Lehmann and Stein (1949) and Coudin and Dufour (2009). Sign-based methods provide valid results under very few assumptions. Especially, they allow for general forms of nonlinear dependence in the data. For example, the shape of the error distribution may depend on the instruments provided a sign invariance condition is satisfied. Our approach, which can be applied in time series and in cross-section contexts, extends that part of the literature.[1]

Other test procedures, which are valid in the presence of weak instruments, are parametric or asymptotically justified. A first approach exploits AR-type statistics; see Dufour (1997), Dufour and Jasiak (2001) and Stock and Wright (2000). More recently, Dufour and Taamouti (2005) extended the AR procedure to construct a whole system of inference on the structural parameters. They derived closed-form solutions for the simultaneous confidence regions and for projection-based confidence intervals in special cases. The second approach, followed by Kleibergen (2002, 2005, 2007), considered a score-type statistic in the limited information simultaneous equation model (LISEM). The so-called K statistic, which is asymptotically a pivotal function, does not depend on the number of instruments, in contrast with AR tests which loose power when many instruments are involved in the model. In a Gaussian context, Bekker and Kleibergen (2003) investigated the K statistic properties in finite samples. They derived a conservative inference by bounding its behavior. Finally, the conditional approach proposed by Moreira (2003) relies on similar tests; see also Moreira (2001), Moreira and Poi (2003), Cruz and Moreira (2005), Andrews, Moreira, and Stock (2004, 2007). Under the null hypothesis, the size of similar tests does not depend on unknown parameters (especially the endogenous explanatory variables and the instruments). Consequently, a similar test remains valid in the presence of weak instruments. Moreira showed that similar tests can be constructed from non-similar ones by associating a critical value function of those unknown parameters. The conditional likelihood ratio test (CLR) so derived exhibits the best properties. Heteroskedastic and autocorrelation corrected versions of the K and the LR statistics are proposed by Kleibergen (2007). See also Andrews and Stock (2005) for a complete review of the IV literature.

The sign-based approach is in the spirit of Anderson and Rubin.[2] Basically, test statistics are obtained by regressing the signs of the constrained residuals on auxiliary regressors (the instruments) with the particularity that tests are performed using the *exact* distribution of those statistics. Like

[1] For a recent econometric exploitation of this sign invariance property, see also Chernozhukov, Hansen, and Jansson (2009) who proposes a finite-sample testing approach in a generic quantile model with conditionally independent sampling.

[2] It is also related to Moreira's approach since the derived tests are similar.

the AR procedure, a sign-based test may suffer from underrejection when many instruments are involved. This well-known drawback of AR-type procedures is corrected by considering "optimal" instruments which maximize test power. Two optimality concepts are considered: the first one leads to locally optimal tests in the neighborhood of the tested value; the second one to point-optimal tests against a particular alternative. Approximate optimal instruments are constructed by split-sample methods; see Angrist and Krueger (1995), Dufour and Jasiak (2001), Dufour and Taamouti (2005).

Other works on median (and quantile) regression with endogenous regressors have focussed on estimation. The starting point was the two-stage-least-absolute-deviation estimator (2SLAD) introduced by Amemiya (1982), which is an adaptation of 2SLS to the least absolute value (LAV) regression [see also Powell (1983) for the asymptotic properties]. In a first stage, the endogenous variable is regressed by ordinary least squares on the instruments. The second stage consists in a LAV regression which involves the fitted values of the endogenous variable. Chen and Portnoy (1996) extended the idea of two step-estimation to other quantiles. Two robust IV quantile estimators based on GMM formulations are due to Honore and Hu (2004). The first one involves signs of the residuals and the second one their ranks. In a linear median regression model, Hong and Tamer (2003) proposed a minimum distance kernel-based estimator that can be used both in a point identified setup or when there exists a set of observationally equivalent parameters. Besides, control function approaches were used by Lee (2003) in a partially linear quantile regression, by Chernozhukov and Hansen (2008) with a double simultaneous optimization,[3] and by Sakata (2001) who proposed a general approach also based on a double optimization of the ratio between the error dispersion controlled by the instruments and the dispersion without control. Here, we propose to associate a Hodges-Lehmann-type estimator to the finite-sample-based inference results when the parameter is identified. The estimate (or the set of estimates) is the (set of) value(s) least rejected by sign-based tests, or equivalently the one(s) leading to the highest p-value [see Hodges and Lehmann (1963) and Coudin and Dufour (2008)].

The paper is organized as follows. The model and notations are presented in section 2. In section 3, general results on the finite-sample sign-based inference are stated: the distribution of the constrained signs is derived under the sign invariance assumption. Then, simultaneous tests with controlled level are constructed by Monte Carlo test techniques. Further, confidence sets and general tests are built using projection techniques. In sections 4 and 5, we go further in details and choose the form of the sign-based test statistics on the basis of power properties. Pointwise and local optimality concepts are both considered for choosing the instruments. We also follow two different approaches for determining the form of the sign-based statistic. First, we study a classical GMM statistic that is a quadratic form of the residual signs with a certain weight matrix. We also consider a Tippett-type combination [Tippett (1931)], which relies on the minimum of the p-values corresponding to each sign-based moment equation tested separately. Section 6 is dedicated

[3]Their estimate of the parameter suffering from endogeneity both satisfies the regression criterion minimization and minimizes the instrumental regressors parameters norm. They also obtain valid confidence regions by test inversion.

to asymptotic properties of the proposed test procedures under assumptions weaker than the ones required for finite-sample validity. Section 7 presents IV sign-based estimators when identification holds. The power performances of the sign-based methods are compared to other usual methods in the simulation studies of section 8. Finally, an illustrative application to the returns to schooling [Angrist and Krueger (1991)] is provided in section 9. We conclude in section 10. Appendix A contains the proofs.

2. Framework

In this section, we extend the linear median regression framework used in Coudin and Dufour (2009) and Coudin and Dufour (2008) to a nonlinear and instrumental setup. Let $\{W_t = (y_t,\ x_t',\ z_t') : \Omega \to \mathbb{R}^{p+k+1}\}_{t=1,\dots,n}$ be a stochastic process defined on a probability space $(\Omega, \mathcal{F}, P)$, and $\{W_t, \mathcal{F}_t\}_{t=1,\dots,n}$ an adapted stochastic sequence where $\mathcal{F}_t$ is a σ-field in Ω such that $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s < t$ and $\sigma(W_1, \dots, W_t) \subset \mathcal{F}_t$. y_t is the real dependent variable, which can take continuous or discrete values, $x_t = (x_{t1}, \dots, x_{tp})'$ is a p-vector of explanatory variables (possibly endogenous) and $z_t = (z_{t1}, \dots, z_{tk})'$ is a k-vector of exogenous variables. We further assume that y_t, x_t and the parameter of interest, $\theta \in \mathbb{R}^q$, are related through a nonlinear function $f :\ \mathbb{R}^{1+p+q} \to \mathbb{R}$ up to an error term u_t:

$$f(y_t, x_t, \theta) = u_t,\ t = 1, \dots, n.$$

For convenience, we will use the following matrix notation

$$f(y, X, \theta) = u \tag{2.1}$$

where $y = (y_1, \dots, y_n)'$ and $u = (u_1, \dots, u_n)'$ are real n-vectors, $X = (x_1, \dots, x_n)'$ is a $n \times p$ real matrix.

We denote $Z = (z_1, \dots, z_n)'$ the $n \times k$ real matrix of instruments. The terminology of instruments is very general. It covers exogenous random variables but the instruments may also depend on the parameter θ such as a score vector in a nonlinear model. In such a case, we shall denote $Z_\theta := \big(z_1(\theta), \dots, z_n(\theta)\big)'$. Instruments may be strongly or weakly correlated with the endogenous regressors, but they have to be valid in the following sense.

Assumption A1 Z-CONDITIONAL MEDIANGALE. *Let $\{u_t, \mathcal{F}_t\}_{\ t=1,2,\dots}$ be an adapted stochastic sequence and $\mathcal{F}_t = \sigma(u_1, \dots, u_t, Z)$. We assume that*

$$P[u_1 > 0|Z] = P[u_1 < 0|Z] = 1/2,$$
$$P[u_t > 0|Z, u_{t-1}, \dots, u_1] = P[u_t < 0|Z, u_{t-1}, \dots, u_1] = 1/2, \textit{ for } t > 1.$$

Assumption A1 is an adaptation of the mediangale concept defined in Coudin and Dufour (2009) to an instrumental setup. We condition on Z instead of X since some explanatory variables are endogenous. $\{u_t\}_{t=1,\ldots,n}$ are not supposed to be $i.i.d.$. The past values of u_t may have an influence on the form of the distribution of the current u_t, provided they do not affect its probability of being positive or negative. This flexible setup covers the standard limited information simultaneous equations model (LISEM) [see Hausman (1983)]:

$$
\begin{gathered}
y_t = x_t'\theta + u_t, \\
x_t = z_t'\Pi + v_t, \\
\begin{pmatrix} u_t \\ v_t \end{pmatrix} \overset{iid}{\sim} \mathcal{N}(0, \Sigma), \text{ for } t = 1, \ldots, n, \\
(u_t, v_t') \text{ independent of } z_t, \text{ for } t = 1, \ldots, n,
\end{gathered}
$$

where y_t is a scalar dependent variable, x_t is a p-vector of explanatory and possibly endogenous variables, z_t is a k-vector of exogenous variables, u_t is the error term of the structural equation , and v_t is the p-vector of disturbances of the instrumental equation. θ is a p-vector of structural parameters and Π is the $k \times p$ matrix of the reduced form parameters. In a standard LISEM, (u_t, v_t') are $i.i.d.$ normally distributed and independent of z_t.

Model (2.1) with the Assumption A1 is much more general. Parametric assumptions on the error term distribution are relaxed. The normality restriction is not required neither in finite samples nor asymptotically. Assumption A1 allows for heteroskedasticity of unknown form. Only the median is assumed to be zero (conditional on Z). This leads to three important special cases.

First, the independence assumption between the observations is relaxed. Past realizations of u_t can have an influence on the shape of the current u_t distribution. For example, $u_t,\ t = 1, \ldots, n$, can satisfy the following assumptions:

$$
\begin{gathered}
u_1 = \sigma_1 \varepsilon_1 \,, \\
u_t = \sigma_t(u_1, \ldots, u_{t-1})\varepsilon_t \,, \quad \text{for } t = 2, \ldots, n \\
\varepsilon_1, \ldots, \varepsilon_n \text{ are independent with median zero,} \\
\sigma_1 \text{ and } \{\sigma_t(u_1, \ldots, u_{t-1})\}_{t=2,\ldots,n} \text{ are non-zero with probability one.}
\end{gathered}
\tag{2.2}
$$

This includes in a time series context ARCH(q) with non-Gaussian noise ε_t, where

$$
\sigma_t(u_1, \ldots, u_{t-1})^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \cdots + \alpha_q u_{t-q}^2. \tag{2.3}
$$

Second, the instruments may have an influence on the shape of the current u_t distribution, provided the probability of being positive or negative is not affected. In finite samples, an instrument affecting the shape of the disturbance distribution, may be the cause of asymptotic test great distortions. Examples can be found in section 8. In such a case, one can exploit Assumption A1 that allows for some nonlinear dependence between Z and u, for any sample size. A large spectrum of heteroskedastic patterns is covered, such as:

$$u_t = \sigma_t(Z)\,\varepsilon_t,\ t = 1, \ldots, n, \tag{2.4}$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are $i.i.d.$ conditional on Z. This can be useful when the instrument choice is limited by data availability.

A third interesting case arises when the endogenous variables affect the shape of the structural error distribution. The usual linear specification simplifies calculus and interpretation. However, if the relation is not well captured by linear modeling, the shape of the structural error distribution may be affected. In such a case, asymptotic tests are invalid even in a large sample.

When u_t and z_t are only asymptotically uncorrelated, Assumption A1 may not hold ($e.g.$ due to feedback on the error signs). However, we will see below that sign-based tests are still asymptotically valid.

3. Finite-sample inference with possibly weak instruments

Assumption A1 is the cornerstone of the validity of sign-based inference methods. If the disturbances satisfy a conditional mediangale condition, their signs have a known joint distribution that does not depend on any nuisance parameter (conditional on the instruments). This property holds for any sample size, without imposing additional distributional assumptions. The sign pivotality property was stated in Coudin and Dufour (2009) for classical median regressions. It was exploited to construct sign-based simultaneous tests with controlled level for any sample size by Monte Carlo test techniques. In that section, we extend that result to nonlinear and possibly instrumental regressions. Then, we follow the same strategy and conduct simultaneous tests. More generally, the whole finite-sample based inference system presented in Coudin and Dufour (2009, 2008) applies here. Simultaneous confidence regions with controlled level are constructed by inverting simultaneous tests; and more general confidence sets or tests, by projecting the simultaneous confidence regions. We rapidly present the leading ideas and principles of finite-sample based inference system. For a detailed presentation, the reader is referred to Coudin and Dufour (2009, 2008).

3.1. Pivotality

Let us begin with some notations. We define the sign operator $s : \mathbb{R} \to \{-1, 0, 1\}$ as

$$s(a) = \mathbf{1}_{[0,+\infty)}(a) - \mathbf{1}_{(-\infty,0]}(a),\ \text{where}\ \ \mathbf{1}_A(a) = \begin{cases} 1,\ \text{if}\ a \in A, \\ 0,\ \text{if}\ a \notin A. \end{cases} \tag{3.1}$$

For convenience, the notation will be extended to vectors. Let $u \in \mathbb{R}^n$ and $s(u)$, the n-vector composed by the signs of its components. This enables us to formally state the following proposition:

Proposition 3.1 SIGN DISTRIBUTION. *Under model* (2.1)*, suppose the errors* $(u_1, \ldots, u_n)$ *satisfy Assumption A1 conditional on* Z_θ*. Then the variables* $s(u_1), \ldots, s(u_n)$ *are i.i.d. conditional on* Z_θ *according to the distribution*

$$\mathsf{P}_\theta[s(u_t) = 1|Z_\theta] = \mathsf{P}_\theta[s(u_t) = -1|Z_\theta] = 1/2\,, \quad t = 1, \ldots, n\,. \tag{3.2}$$

The proofs of the theorems and propositions appear in the Appendix.
From the latter proposition, it follows that the vector of constrained signs

$$s(f(y, X, \theta)) := \big(s(f(y_1, x_1, \theta)), \ldots, s(f(y_n, x_n, \theta))\big)' \tag{3.3}$$

has a nuisance-parameter-free distribution (conditional on Z), $i.e.$ it is a **pivotal function**. When the disturbance process satisfies Assumption A1, the error signs are mutually independent according to a known distribution.
Furthermore, any real-valued function of the form

$$\overline{T}_\theta\big(y, \theta\big) = T\big(s(f(y, X, \theta)),\, Z_\theta, \theta\big) \tag{3.4}$$

has a distribution which does not depend on unknown nuisance parameters. Its conditional distribution given Z_θ can be analytically derived or simulated because the joint distribution of $s(f(y, X, \theta))$ is completely specified by Proposition **3.1**. Consequently, we can construct conditional tests for which size is fully controlled.
Consider the problem of testing

$$H_0(\theta_0):\ \theta = \theta_0 \text{ vs } H_1(\theta_0) : \theta \neq \theta_0.$$

Under H_0,

$$T\big(s(f(y, X, \theta_0)), Z_{\theta_0}, \theta_0\big) \sim T(S_n, Z_{\theta_0}, \theta_0) \tag{3.5}$$

where $S_n = (s_1, \ldots, s_n)'$ and $s_1, \ldots, s_n$ are $i.i.d.$ Bernoulli random variables conditional on Z_{θ_0} that equal 1 with probability $1/2$ and -1 with probability $1/2$. A test with level α rejects the null hypothesis when

$$T\big(s(f(y, X, \theta_0)), Z_{\theta_0}, \theta_0\big) > c_T(Z_{\theta_0}, \alpha, \theta_0) \tag{3.6}$$

where $c_T(Z_{\theta_0}, \alpha, \theta_0)$ is the $(1-\alpha)$-quantile of the distribution of $T(S_n, Z_{\theta_0}, \theta_0)$ conditional on Z_{θ_0}.

This property is an extension of the one stated in Coudin and Dufour (2009); see also Dufour (1981), Campbell and Dufour (1991, 1995) and Wright (2000).[4] Here, $T\big(s(f(y, X, \theta_0)), Z_{\theta_0}, \theta_0\big)$

[4] For an econometric exploitation of a version of the sign invariance property adapted to generic quantile regression with independent conditional sampling, see also Chernozhukov, Hansen, and Jansson (2009).

and Z_{θ_0} depend on the tested value θ_0. This property can be adapted to error distributions with a mass at zero by randomly breaking the zeros in a way similar to Coudin and Dufour (2009).

Furthermore, the sign pivotality result allows one to construct nonparametric tests through Monte Carlo test techniques.

3.2. Monte Carlo tests

Under $H_0(\theta_0)$ and Assumption A1, the conditional distribution of $T_{\theta_0}\big(s(f(y, X, \theta_0)), Z_{\theta_0}\big)$ given Z_{θ_0} is free of nuisance parameters with a known distribution that can be simulated. Those two features are sufficient to apply Monte Carlo test procedures.[5] Given T_{θ_0}, the test proposed in section 2 rejects $H_0(\theta_0)$ when $T_{\theta_0} \geq c$, with c depending on the level. The general idea of Monte Carlo tests is to order the observed statistic with N simulated ones. The Monte Carlo test rejects $H_0(\theta_0)$ when the observed statistic is larger than at least $(1-\alpha)\times N$ simulated replicates. As the distribution of T_{θ_0} is discrete, we need a criterion to order two equal realizations. We shall use the randomized tie-breaking presented in Dufour (2006) and Coudin and Dufour (2009).

The Monte Carlo test for $H_0(\theta_0)$ can equivalently be conducted with empirical p-values. Let $T^{(0)}_{\theta_0}$ be the "observed" statistic, $(T^{(1)}_{\theta_0}, \ldots, T^{(N)}_{\theta_0})$ be a N-vector of independent replicates drawn from the same distribution as T_{θ_0}, and $(W^{(0)}, \ldots, W^{(N)})$ be a $N+1$-vector of $i.i.d.$ real uniform variables. A Monte Carlo test with level α consists in rejecting the null hypothesis whenever the empirical p-value, denoted $\tilde{p}^{\theta_0}_N(T^{(0)}_{\theta_0})$, is smaller than α with

$$\tilde{p}^{\theta_0}_N(x) = \frac{N\tilde{G}^{\theta_0}_N(x)+1}{N+1}, \tag{3.7}$$

where

$$\tilde{G}^{\theta_0}_N(x) = 1 - \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}_{[0,\infty)}(x - T^{(i)}_{\theta_0}) + \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}_{[0]}(T^{(i)}_{\theta_0} - x)\mathbf{1}_{[0,\infty)}(W^{(i)} - W^{(0)})$$

is the simulated survival function. If N is such that $\alpha(N+1)$ is an integer

$$P[\tilde{p}^{\theta_0}_N(T^{(0)}_{\theta_0}) \leq \alpha] = \alpha \text{ for } 0 \leq \alpha \leq 1.$$

The Monte Carlo test so obtained has size α for any given sample size T. No identification condition is needed to conduct tests with fully controlled level. The instruments may be poorly informative, the test levels are always controlled provided that the instruments are exogenous in the sense of Assumption A1. We shall see later on that Assumption A1 can be slightly relaxed while maintaining the test levels *asymptotically* controlled.

[5] See Dwass (1957), Barnard (1963) and Dufour (2006)

Those basic joint tests constitute the matrix for a whole nonparametric inference system where simultaneous confidence regions are obtained by test inversion and tests of general hypothesis by projection techniques.

3.3. Confidence sets, projection-based confidence intervals and confidence distributions

We use the simultaneous sign-based tests to build confidence sets for θ with given level. These are obtained in the following way: Monte Carlo sign-based tests for $H_0(\theta_0)$ are performed for any value of $\theta_0 \in \mathbb{R}^q$ (or more reasonably for a grid of values) yielding a p-value $\tilde{p}_N^{\theta_0}(T_{\theta_0}^{(0)})$. This associated p-value reflects the *degree of confidence* one may have in the hypothesis $\theta = \theta_0$ given the realization $T_{\theta_0}^{(0)}$ [see Coudin and Dufour (2008)]. The simultaneous confidence region with level $1 - \alpha$ is composed by the values of θ_0 with p-value higher than α. Next, from this simultaneous confidence set for θ, it is possible to derive confidence intervals for the individual components and to perform tests for general nonlinear hypotheses using projection techniques.[6] In Coudin and Dufour (2008), we directly applied projection techniques on the simulated p-value function. The projected p-value function associated with the individual component θ_k gives a graphical summary of the inference results on θ_k.

The functions involved here are highly nonlinear and no closed-form analytical solutions can easily be obtained. Practical implementation requires to solve optimization problems under non-linear constraints. Search programs such as simulated annealing are used [see Goffe, Ferrier, and Rogers (1994) and Press, Teukolsky, Vetterling, and Flannery (2002)].

3.4. Simplifications: restrictions on the parameter space

This approach requires in theory to evaluate the sign-based statistic for any value of the parameter in the parameter space. When the size of the parameter space increases, the search programs rapidly become computationally intensive especially when projection techniques are used. So, any additional piece of information that helps to reduce the size of the parameter space is welcome and must be included as a constraint in the program. First of all, restrictions implied by the economic theory or by the relevance of the model have to be taken into account. If the underlying economic model specifies that a certain coefficient must be less than one (such as an elasticity for example), there is no use to investigate what happens outside.

More generally, a conditional approach is also possible. If one accepts to fix some of the parameter components in a certain subspace, say Θ^c, the approach presented above gives results conditional on θ belonging to Θ^c.

[6] For examples in different settings and for further discussion on projection techniques, the reader is referred to Coudin and Dufour (2009), Dufour (1990), Dufour (1997), Wang and Zivot (1998), Dufour and Jasiak (2001), Dufour and Taamouti (2005).

An alternative approach consists in restricting the parameter space to a consistent set estimator. Such confidence-set restricted Monte-Carlo tests are asymptotically valid under some general regularity conditions; see Dufour (2006).

The two following sections are dedicated to the construction of efficient test statistics which satisfy the general form $T_\theta\big(s(f(y, X, \theta)), Z_\theta\big)$ so that the finite-sample inference system can be applied. We consider two approaches. First, we establish the general form of point-optimal tests versus a specified alternative. This theoretical result yields a power frontier for sign-based procedures. However, methods that combine various point-optimal tests to approach the power envelope are not easily tractable in practice. Hence, we turn to a more classical approach and derive locally optimal instruments. We study statistics that involve signs in a quadratic form and a Tippett-type combination although other (less usual) statistics could also be envisaged ($e.g.$ linear plus quadratic forms or polynomials at various orders involving signs). The class of quadratic IV-type sign-based statistics provides good competitors when the final aim is estimation.

4. Point-optimal tests

Point-optimal tests are usually derived for parametric models since they rely on the likelihood ratio that follows from the classical Neyman-Pearson lemma. Here, they can be constructed for nonparametric models thanks to the sign transformation. In this section, we present point-optimal tests for signs in a general context and then, in a regression context.

4.1. General point-optimal sign-based result

Point-optimal tests based on signs are derived for a very general nonparametric framework in which signs are independent and heterogeneously distributed according to Bernoulli distributions with parameters $(p_1, ..., p_n)$.

$$P[s_t = 1] = p_t, \quad P[s_t = -1] = 1 - p_t, \quad t = 1, \ldots, n. \tag{4.8}$$

Let us consider the problem of testing

$$H_0: \quad (p_1, ..., p_n)' = (p_{01}, ..., p_{0n})', \tag{4.9}$$

versus

$$H_1: \quad (p_1, ..., p_n)' = (p_{11}, ..., p_{1n})'. \tag{4.10}$$

Proposition 4.1 POINT-OPTIMAL SIGN-BASED TEST. *When testing H_0 versus H_1, the most* *powerful test based on signs rejects H_0 when*

$$\sum_{t=1}^{n} s_t \ln\left(\frac{p_{1t}(1-p_{0t})}{p_{0t}(1-p_{1t})}\right) > c(\alpha, H_1)$$

with $c(\alpha, H_1)$ depending on the level.

The proof is a direct application of the Neyman-Pearson lemma [see for example Gouriéroux and Monfort (1995)]. Point-optimal tests are often derived in parametric setups because they rely on the form of the likelihood function under the null hypothesis and under the alternative. Here, the point-optimal test can be derived in a nonparametric setup thanks to the sign transformation. The main strength of the sign transformation is indeed to get rid of the distributional characteristics of the underlying process. However, one has to choose the alternative hypothesis to specify $\{p_{1t}\}_{t=1,\dots,n}$.

4.2. Point-optimal sign-based tests in a regression framework

We now go back to the regression framework of model (2.1) with Assumption A1. Consider testing $H_0 :\ \theta = \theta_0$ versus $H_1 :\ \theta = \theta_1$, Proposition **4.1** yields the following corollary.

Corollary 4.2 POINT-OPTIMAL SIGN-BASED TEST IN A REGRESSION CONTEXT. *In model* (2.1)*, let $\{W_t = (y_t,\ x_t',\ z_t')\}_{t=1,\dots,n}$ be a i.i.d. process and $\{u_t\}_{t=1,\dots,n}$ have a common distribution function G conditional on Z that does not depend on θ. Suppose further that the mediangale Assumption A1 holds. Then the most powerful sign-based test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ in the sense of Neyman-Pearson rejects H_0 when*

$$\sum_{t=1}^{n} s(u_t) \ln\left(\frac{1-G(h_t)}{G(h_t)}\right) > c(\alpha, \theta_1) \tag{4.11}$$

where $(h_1, \dots, h_n)' = (f(y_1, x_1, \theta_1) - f(y_1, x_1, \theta_0), \dots, f(y_n, x_n, \theta_1) - f(y_n, x_n, \theta_0))'$, and $c(\alpha, \theta_1)$ depends on the level.

The point-optimal sign-based test is a linear form of the signs with weights depending on the error distribution and the chosen alternative hypothesis. When the distribution function G is logistic, the statistic simplifies and the optimal weights turn out to be $\{h_t\}_{t=1,\dots,n}$.

Point-optimal sign-based tests are theoretically interesting objects because they bound what can be done with signs and combining them allows one to approach the power envelope. However, a point-optimal test requires first to specify the alternative hypothesis and then to compute the optimal weights $\{p_{1t}\}_{t=1,\dots,n}$ that depend on the error distribution. In a parametric setup, this can be done analytically. But in a nonparametric setup (as here), the error distribution is not fixed and $\{p_{1t}\}_{t=1,\dots,n}$ are not straightforward to choose. Point-optimal statistic can be approached if one "guesses" the behavior of the error term under the alternative hypothesis. This can be done by split-sample techniques. A first part of the sample is used to approach the error distribution, the other part, to construct the statistic; see Dufour and Taamouti (2010) for an example of use.

However, approaching point-optimal tests and power envelope quickly become computationally intensive. For this reason, we turn in the next section to another optimality concept that does not require to specify the alternative hypothesis and still provides "locally" optimal tests. The so-called locally optimal test statistics turn to be quadratic forms of the constrained signs and of optimal instruments. We also study other combinations (than quadratic) of sign-based moment equations that may present power in weak identified cases.

5. IV sign-based statistics

The easiest way to introduce IV sign-based statistics is to refer to a GMM setup. Signs and instruments that satisfy the mediangale Assumption A1 also satisfy usual moment conditions. GMM statistics exploiting the orthogonality between the error signs and the instruments can be constructed using the analogy principle. More generally, we follow the idea of auxiliary regressions [Anderson and Rubin (1949) and Dufour (2003)] to circumvent the problem of endogeneity; see also the artificial regressions of Davidson and McKinnon (2001). We consider regressions of the constrained signs on "auxiliary" instruments (when present in the model their coefficient must be zero). We consider two approaches. IV sign-based statistics correspond either to F-type statistics for testing that the parameter vector in the previous multivariate regression is zero (denoted GMM-type), either to Tippett-type combination of univariate regressions involving one "auxiliary" regressor at once (denoted Tippett-type). The proposed sign-based statistics are pivotal functions and exact sign-based tests can be built for any sample size regardless of the strength of the instruments. Then, we focus on IV sign-based statistics that yield to the best (local) power considerations and on the corresponding optimal instruments.

5.1. Sign-based moment equations

In a usual LISEM model (with valid instruments), the estimating equations correspond to the orthogonality conditions between z_t and u_t.

$$E[(y_t - x_t\theta)z_{jt}] = 0, \quad \text{for } j = 1, \dots, k,\ t = 1, \dots, n. \tag{5.12}$$

Under Assumption A1, Proposition **3.1** entails that the error signs are $i.i.d.$ conditional on Z and centered. Consequently, in model (2.1), the following "sign-based" moment conditions (where the residuals are replaced by their signs) hold:

$$E\big[s\big(f(y_t, x_t, \theta)\big)z_{jt}\big] = 0, \quad \text{for } j = 1, \dots, k,\ t = 1, \dots, n. \tag{5.13}$$

More generally, Assumption A1 entails

$$E\left\{s\big(f(y_t, x_t, \theta)\big)g_j\big(z_t(\theta), \theta\big)\right\} = 0, \quad \text{for } j = 1, \dots, J,\ t = 1, \dots, n. \tag{5.14}$$

where $\{g_j\}_{j=1,\dots,J}$ are measurable functions of the instruments and θ.[7] If necessary, we shall redefine instruments as $\tilde{z}_{jt}(\theta) = g_j\big(z_t(\theta), \theta\big),\ t = 1, \dots, n,\ j = 1, \dots, J$ but the following applies without any further modification.
In those sign-based moment equations, the parameter of interest is not present in an explicit form but is implicitly involved through a robust transformation by the sign operator. The sign operator gets rid of any nuisance parameter affecting the distribution of the error term and enables one to conduct fully robust tests to heteroskedasticity of unknown form for any sample size.
The analogy principle entails the following sample-based moment equations:

$$\sum_{t=1}^{n} s\big(f(y_t, x_t, \theta)\big) z_{jt} = 0,\ j = 1, \dots, k. \tag{5.15}$$

5.2. Combining sign-based moment equations: GMM or multiple tests

These new orthogonality conditions can be exploited for constructing GMM-type statistics. For testing $H_0(\theta_0) : \theta = \theta_0$ versus $H_1(\theta_0) : \theta \neq \theta_0$ in model (2.1), we shall consider test statistics of the following form:

$$D_S(\theta_0, Z, \Omega_n) = s(f(y, X, \theta_0))' Z_{\theta_0} \Omega_n\big(s(f(y, X, \theta_0)), Z_{\theta_0}\big) Z'_{\theta_0} s(f(y, X, \theta_0)) \tag{5.16}$$

where $\Omega_n\big(s(f(y, X, \theta_0)), Z_{\theta_0}\big)$ is a $k \times k$ positive definite weight matrix that may depend on the constrained signs $s(f(y, X, \theta_0))$ under $H_0(\theta_0)$.
The statistic associated with $\Omega_n = (Z'_{\theta_0} Z_{\theta_0})^{-1}$ is given by: [8]

$$D_S\big(\theta_0, Z_{\theta_0}, (Z'_{\theta_0} Z_{\theta_0})^{-1}\big) = s\big(f(y, X, \theta_0)\big)' P(Z_{\theta_0}) s\big(f(y, X, \theta_0)\big) \tag{5.17}$$

where $P_{Z_{\theta_0}} = Z_{\theta_0} (Z'_{\theta_0} Z_{\theta_0})^{-1} Z'_{\theta_0}$. That is the squared norm of the fitted values from the regression of $s(f(y, X, \theta_0))$ on Z_{θ_0}. In other words, $D_S\big(\theta_0, Z_{\theta_0}, (Z'_{\theta_0} Z_{\theta_0})^{-1}\big)$ is a monotonic transformation of the Fisher statistic for testing $\gamma = 0$ in the artificial regression model $s\big(f(y, X, \theta_0)\big) = Z_{\theta_0}\gamma + v$.

Another way to approach the problem of building sign-based statistics is then to consider regressions of the constrained signs on appropriately chosen “instruments”:

$$s\big(f(y, X, \theta_0)\big) = Z_{\theta_0}\gamma + v\ . \tag{5.18}$$

Testing $H_0(\theta_0)$ is equivalent to test $\gamma = 0$ in (5.18) where $\tilde{Z}(\theta_0)$ are related to X but excluded from the structural model. $\tilde{Z}(\theta_0)$ are called "auxiliary regressors": when present in the model, their coefficient must be zero. Remark that the unilateral point-optimal test presented in Proposition **4.1** can also be viewed as a t-test obtained by regressing the signs on some appropriate auxiliary

[7] Hong and Tamer (2003) proposed for example to use kernel functions.

[8] This is the GMM statistic studied by Chernozhukov, Hansen, and Jansson (2009) in their conditionally independent setting.

instruments (precisely the scores under the alternative). Thus, the set of test-statistics based on auxiliary instruments is very general and includes point-optimal Neyman-Pearson-type statistics among the related t-statistics.

Fisher and GMM-type statistics are quadratic forms of the moment equations. Other types of combination of sign-based moment equations can be exploited. We can for example follow Tippett (1931) and consider

$$D_S^{Tip}(\theta_0, Z_{\theta_0}) = min(p_1, \ldots, p_k) \tag{5.19}$$

where $p_1, \ldots, p_k$ are the (empirical) p-values associated with testing $\gamma_i = 0$ in the univariate regression involving one instrument (here $z_{i\theta_0}$) at once:

$$s(f(y, X, \theta_0)) = \gamma_i z_{\theta_0 i}, \; i = 1, \ldots, k. \tag{5.20}$$

The idea behind is the following. Statistics based on a quadratic combination of moment equations are specifically adapted for test and estimation when the parameter is well identified because they rely a local optimality concept. However, in weakly identified cases, there is no gain to restrict on statistics that provide power in the vecinity of the true value parameter because those values may be observationally equivalent (due to the lack of identification). In such cases, other combinations of the moment equations such as the Tippett combination may provide better overall properties.

5.3. Artificial regressions

The use of artificial regressions such as (5.18) and (5.20) to circumvent endogeneity has been first proposed by Anderson and Rubin (1949) [see also Dufour (2003), Davidson and McKinnon (2001) who presented artificial regressions in general nonlinear models]. In the linear Gaussian model, they proposed an exact test of $\gamma = 0$ based on a Fisher-type statistic. The derived inference is valid and robust to possibly weak instrument settings [see also Dufour (1997), Staiger and Stock (1997), Dufour and Taamouti (2005)]. However, the procedure power depends on the choice of the instruments. In the LISEM model with exact identification and Gaussian disturbances the AR procedure is optimal, but it may suffer from underrejection when a large number of instruments is involved in the model. With "many instruments", asymptotically justified methods such as Kleibergen's K statistic or Moreira's LM statistic may provide better asymptotic power. However, those statistics are no longer pivots in finite samples and a relying inference without other adjustment may suffer from size distortion even in a Gaussian context.[9]

Here, our objective is double. We propose test statistics that are first pivotal functions for any sample size, under the null hypothesis and with known distribution, in order to conduct exact inference ($i.e.$ that satisfy Assumption A1) and that are based on an "optimal" choice of instruments.

[9]The K statistic distribution depends on nuisance parameters in finite samples. In a Gaussian context, Bekker and Kleibergen (2003) derived bounding distributions and conservative tests.

5.4. Locally optimal instruments

In case of overidentification, instruments can be selected to improve power consideration. When testing $H_0(\theta_0)$ with level α, the power function of the sign-based statistics $T\big(s(f(y,X,\theta_0)), Z_{\theta_0}\big)$ is:

$$\beta(\theta) = P_\theta\big[T\big(s(f(y,X,\theta_0)), Z_{\theta_0}\big) > c_T(Z_{\theta_0}, \alpha)\big]. \tag{5.21}$$

We search for instruments that "maximize" the power function locally around θ_0 in a just identified setup.[10] Around θ_0, sign-based test power functions follow the behavior of their second derivatives $w.r.t.$ θ, which turn to be quadratic forms of the sign vector. Consequently, we derive the optimal instruments from the weights involved in the latter quadratic forms and derive locally optimal sign-based test statistics. This result is stated in the following proposition. Locally optimal instruments are derived in a setup with $i.i.d.$ observations. In the sequel, all results are conditional on the available set of instruments.

Proposition 5.1 LOCALLY OPTIMAL INSTRUMENTS. *Consider the problem of testing $H_0: \theta = \theta_0$, in model (2.1) versus a sequence of alternatives $H_n: \theta = \theta_n$ such that $\theta_n \underset{\theta_n \neq \theta_0}{\rightarrow} \theta_0$, and assume that:*

- **a)** $(y_t, x_t, z_t),\ t = 1, \dots, n$ *are identically and continuously distributed;*
- **b)** f *is continuously differentiable in θ, with continuous derivative $H_t(\theta) = \frac{\partial f(y_t,x_t,\theta)}{\partial \theta'}\Big|_\theta$ and $H(\theta)' = (H_0(\theta)', \dots, H_n(\theta)')'$ for $t = 1, \dots, n$;*
- **c)** $\exists V(\theta_0)$ *such that*

$$\sup_{\theta \in V(\theta_0)} \left|\left| E\left[\frac{\partial f(y_t,x_t,\theta)}{\partial \theta'}\bigg|_\theta \right] \right|\right| = \sup_{\theta \in V(\theta_0)} ||E[H_t(\theta)]|| \leq M_1, \quad \forall t = 1, \dots, n;$$

- **d)** u_t *has continuous distribution function G which is continuously differentiable at zero with derivative G' also continuously differentiable at zero and $G''(0) = 0$, for $t = 1, \dots, n$;*
- **e)** *setting* $P_{\theta_n}\big[u_t - \big(H_t(\overline{\theta}) - EH_t(\overline{\theta})\big)(\theta_n - \theta_0) \leq x\big] = G_n^{\overline{\theta}}(x)$,

$$\frac{1}{||\theta_n - \theta_0||}\big(G_n^{\overline{\theta}}(0) - G(0)\big) \to 0 \text{ and } \big(G_n'^{\overline{\theta}}(0) - G'(0)\big) \to 0,$$
$$\text{for all } \overline{\theta} \text{ such that } ||\theta_0 - \overline{\theta}|| \leq ||\theta_0 - \theta_n||.$$

Then, a locally optimal set of instruments is given by

$$Z^*(\theta_0) = E[H(\theta_0)], \tag{5.22}$$

[10] Another alternative is to compute instruments maximizing the power function versus a specified alternative. This strategy has been followed by Dufour and Taamouti (2002) who derived point-optimal AR tests in a Gaussian context.

and a locally optimal GMM sign-based statistic is

$$D_S^*(\theta_0) = s\big(f(y,x,\theta_0)\big)' EH(\theta_0)[EH(\theta_0)' EH(\theta_0)]^{-1} EH(\theta_0)' s\big(f(y,x,\theta_0)\big). \tag{5.23}$$

The regularity conditions **b**,**c** and **d** insure continuity, differentiability and integrability of f and of its derivatives. Condition **d** states that the errors possess a mode at zero. Further, condition **e** sets the speed of convergence of the distribution functions G_n towards G. Further, if $u_t - [H_t(\overline{\theta}) - EH_t(\overline{\theta})](\theta_n - \theta_0)$ has a symmetric distribution for any value of θ_n then condition **e** holds.

If the matrix $H_t(\theta_0)$ is exogenous it can directly be used. If not, we need an exogenous estimate to ensure inference validity for a given n. This is feasible by splitting the sample into two parts.

5.5. Quasi-optimal instruments and split-sample

When observations are independent, one may resort to split-sample techniques.[11] The principle is the following. The sample is divided into two parts: $(Y_{(1)}, X_{(1)}, Z_{(1)})$ and $(Y_{(2)}, X_{(2)}, Z_{(2)})$. The first part is used to estimate

$$\frac{\partial f(Y_{(1)}, X_{(1)}, \theta)}{\partial \theta'}\bigg|_{\theta=\theta_0}' = h(Z_{(1)}, \theta_0) + \epsilon, \tag{5.24}$$

yielding an estimate $\hat{h}$. This first stage regression may be linear or not, parametric or not depending on the structural model. A sign-based estimation can also be used.

Then, quasi-optimal instruments are constructed for the second part of the sample, $\tilde{Z}_{(2)} = \hat{h}(Z_{(2)})$ and used as auxiliary regressors in the second step regression:

$$s(f(Y_{(2)}, X_{(2)}, \theta_0)) = \gamma \tilde{Z}_{(2)} + v_{(2)}. \tag{5.25}$$

A test of $H_0(\theta_0)$ is thus based on a GMM sign-based statistic

$$SSS(\theta_0) = s\big(f(Y_{(2)}, X_{(2)}, \theta_0)\big)' \tilde{Z}_{(2)} [\tilde{Z}_{(2)}' \tilde{Z}_{(2)}]^{-1} \tilde{Z}_{(2)}' s\big(f(Y_{(2)}, X_{(2)}, \theta_0)\big). \tag{5.26}$$

The latter statistic does not depend on nuisance parameters under the null hypothesis because $\tilde{Z}_{(2)}$ is exogenous. Consequently, Monte Carlo tests can be used. This point also validates the use of simulation-based statistics such as a Tippett-type statistic

$$TSS(\theta_0) = \min\{p_1, \dots, p_p\} \tag{5.27}$$

[11] The split-sample technique was used by Dufour and Taamouti (2002) in a quite similar context to ours. They search an exogenous estimate of the point-optimal matrix of instruments, which, in a Gaussian context, allow them to construct exact inference based on generalized AR statistics, [see also Angrist and Krueger (1995), Dufour and Jasiak (2001) for other uses and a discussion on the optimal split of the sample].

where $p_1, \ldots, p_p$ are the empirical p-values for testing $\gamma_i = 0$ in the univariate regressions of the form

$$s(f(Y_{(2)}, X_{(2)}, \theta_0)) = \gamma_i \tilde{z}_{i(2)}, \ i = 1, \ldots, p. \tag{5.28}$$

6. Asymptotic properties

A drawback of the mediangale Assumption A1 is the exclusion of linearly dependent processes even though usual asymptotic inference can still be conducted on them. In Coudin and Dufour (2009), we pointed out that heteroskedasticity and autocorrelation corrected sign-based statistics are asymptotically pivotal functions when signs and explanatory variables are uncorrelated. We also showed that Monte Carlo testing method remained asymptotically valid under weaker distributional assumptions than usual asymptotic Wald tests. In particular, heavy-tailed distributions including infinite variance disturbances were covered. In this section, we show these results apply to IV sign-based statistics without any major modification. We established them for a general nonlinear instrumental regression. A sign HAC-statistic with a weight matrix directly derived from the asymptotic covariance matrix of the signs and the instruments, say $D_S\big(\theta, Z, \frac{1}{n}\hat{J}_n^{-1}(Z)\big)$, turns out to be asymptotically $\chi^2(k)$ distributed under H_0 where k is the number of instruments used.

6.1. Asymptotic behavior of IV GMM sign-statistics

We consider model (2.1) with the following assumptions.

Assumption A2 MIXING. $\{(x_t', z_t'(\theta_0), u_t)\}_{t=1,2,\ldots,}$ *is* α*-mixing of size* $-r/(r-2)$ *with* $r > 2$.[12]

Assumption A3 MOMENT CONDITION. $\mathsf{E}[s(u_t) z_t(\theta_0)] = 0, \ \forall t = 1, \ldots, n, \ \forall n \in \mathbb{N}$.

Assumption A4 BOUNDEDNESS. $z_t(\theta_0) = \big(z_{1t}(\theta_0), \ldots, z_{pt}(\theta_0)\big)'$ *and* $E|z_{ht}(\theta_0)|^r < \varDelta < \infty, \ h = 1, \ldots, k, \ t = 1, \ldots, n, \ \forall n \in \mathbb{N}$.

Assumption A5 NON-SINGULARITY. $J_n^{\theta_0} = \mathrm{var}\left[\frac{1}{\sqrt{n}} \sum_{t=1}^{n} s(u_t) z_t(\theta_0)\right]$ *is uniformly positive definite.*

Assumption A6 CONSISTENT ESTIMATOR. $\varOmega_n^{\theta_0}$ *is symmetric positive definite uniformly over* n *and* $\varOmega_n^{\theta_0} - \frac{1}{n}(J_n^{\theta_0})^{-1} \xrightarrow{p} 0$.

Then we have the following asymptotic distribution.

Theorem 6.1 ASYMPTOTIC DISTRIBUTION OF STATISTIC SHAC. *In model* (2.1)*, with Assumptions A2- A6, we have, under* H_0*,*

$$D_S(\theta_0, Z_{\theta_0}, \varOmega_n^{\theta_0}) \to \chi^2(k).$$

[12] See White (2001) for a definition of α mixing.

Corollary 6.2 *In model* (2.1)*, with the mediangale Assumption A1 and Assumption A4. If* $Z'Z/n$ *is positive definite uniformly over* n *and converges in probability to a definite positive matrix, we have under* H_0*,*

$$D_S\big(\theta_0, Z, (Z'Z)^{-1}\big) \to \chi^2(k).$$

Theorem **6.1** holds for split-sample statistics with $n_2 \to \infty$ and when Z depends on θ (with Z evaluated at θ_0). The proofs are adaptations of Theorem 6.6 and Corollary 6.7 in Coudin and Dufour (2009).

The $\chi^2(k)$ distribution is familiar in instrumental and weak instruments settings. The statistic $k \times AR$ is asymptotically $\chi^2(k)$ distributed [see Anderson and Rubin (1949), Staiger and Stock (1997), Dufour and Jasiak (2001), Dufour and Taamouti (2005)]. This distribution also bounds the LR and LM statistics [see Wang and Zivot (1998)]. However, the $\chi^2(k)$ distribution is directly related to the number of instruments and the use of many instruments (k large) may entail a power loss. This pleads for the K-statistic favor [see Kleibergen (2002)] in setups with normally distributed disturbances or for any statistic whose distribution does not depend of the number of instruments used. When the setup involves more general processes like non-normal of heteroskedastic errors, there is no reason why the power of a K test would be higher than the one of a sign-based test in finite samples. Nevertheless, if one is concerned about the "many instruments" curse, let us underline that sign-based statistics with quasi-optimal instruments are asymptotically $\chi^2(p)$ distributed as the K-statistic, with the advantage of also providing exact inference in finite samples. Only the combination of a joint testing approach with valid instruments entails exact inference for any sample size.

6.2. Asymptotic validity of Monte Carlo tests

Let a test statistic be asymptotically free of nuisance parameters under H_0, with asymptotic distribution F. Monte Carlo tests that rely on replicates possessing the *same asymptotic distribution* F will asymptotically control the level. This result entails that Monte Carlo tests presented in the previous sections "do at least as well as" asymptotic methods when the mediangale Assumption A1 is relaxed and replaced by a classical moment condition (Assumption A3); see Coudin and Dufour (2009). Moreover, those Monte Carlo tests present two considerable advantages over classical asymptotic methods. First, if mediangale Assumption A1 holds, one is sure that the level of Monte Carlo tests is controlled for any sample size. The second advantage comes from the fact that Monte Carlo tests are constructed with replicates based on the *same* sample size. This differs to a classical Monte Carlo test with replicates constructed from the asymptotic distribution. Simulation studies suggest that such Monte Carlo tests perform an implicit sample-size correction [Coudin and Dufour (2009)]. Indeed, for a given sample size, the distribution of the sign statistic may be closer to the one of the replicates than to the (common) asymptotic distribution. Although the use of such Monte

Carlo tests is asymptotically justified, they can be more reliable in small samples than tests based on asymptotic critical values. Under Assumptions A2- A6, testing

$$H_0(\theta_0) : \theta = \theta_0 \text{ versus } H_1(\theta_0) : \theta \neq \theta_0,$$

with the statistic $D_S\big(\theta, Z_{\theta_0}, \hat{J}_n^{-1}(Z_{\theta_0})\big)$ is conducted in the following way:

1. Observe $D_S^{(0)} = D_S\big(\theta_0, Z_{\theta_0}, \hat{J}_n^{-1}(Z_{\theta_0})\big)$. Draw N replicates of the sign vector as if the n observations were independent. The n components of the replicates are thus independent and drawn from a $B(1, .5)$ distribution.
2. Construct $(D_S^{(1)},\ D_S^{(2)}, \ldots, D_S^{(N)})$, the N *pseudo* replicates of $D_S\big(\theta_0, Z_{\theta_0}, (Z'_{\theta_0} Z_{\theta_0})^{-1}\big)$ under the null hypothesis. We call them *pseudo* replicates because they are drawn as if observations were independent.
3. Draw $N+1$ independent replicates $(W^{(0)}, \ldots, W^{(N)})$ from a $\mathcal{U}_{[0,1]}$ distribution and form the couple $(D_S^{(j)},\ W^{(j)})$.
4. Compute $\hat{p}_n^{(N)}(\theta_0)$ using (3.7).
5. The confidence region $\{\theta \in \mathbb{R}^p | \hat{p}_n^{(N)}(\theta) \geq \alpha\}$ level is at least $1 - \alpha$. We reject $\mathcal{H}_0$ if $\hat{p}_n^{(N)}(\theta_0) \leq \alpha$.

In contrast with Wald-type tests based on LIML or GMM estimators which require identification, those asymptotic results lead to valid inference whatever the informative power of the instruments is and for any degree of identification. Finally, moments and density on the u_t process may not exist.

7. IV sign-based estimators

In the previous sections, we have presented simultaneous tests, confidence sets and more general tests based on signs. Estimation is the last step to a complete the inference system. IV sign-based estimators are obtained in a way similar to the one used for the sign-based estimators studied in Coudin and Dufour (2008) in a linear regression without instrument. The estimators maximize the p-value function of the parameter given the form of the IV sign-based statistic and the sample size. They present the *highest confidence degree* based on the chosen IV GMM sign-based statistic. They also turn out (with probability one) to minimize the quadratic function of the signs that is given by the sign-based statistic. Here, we introduce IV sign-based estimators for a general nonlinear possibly instrumental regression. We show, for those general models, that they are consistent with asymptotic normal distribution.[13]

[13]Estimators based on the Tippett-sign statistic could be defined as solutions of a double optimization problem: maximization of the minimal p-value (a sort of Rawls criteria between the moment equations). That question is not addressed in the present paper.

7.1. IV sign-based estimators under point identification

When θ is identified, we can define an IV sign-based estimator as any solution $\hat{\theta}_n(\Omega_n)$ of the problem

$$\min_{\theta \in \mathbb{R}^p} s(f(y, X, \theta))' Z_\theta \Omega_n \left(s\left(f(y, X, \theta)\right), Z_\theta\right) Z_\theta' s(f(y, X, \theta)). \tag{7.29}$$

IV sign-based estimators are analogues of sign-based estimators studied in Coudin and Dufour (2008). These constitute Hodges-Lehmann-type estimators in the sense that they are associated with the highest degree of confidence one may have in a value of θ given the realization of the sample and the choice of the sign-based test statistic $D_S(Z_\theta, \Omega_n, \theta)$ [Hodges and Lehmann (1963)]. The reader is referred to Coudin and Dufour (2008) for a detailed presentation. IV-sign based estimators can also be interpreted as GMM estimators exploiting the orthogonality between error signs and instruments. See Honore and Hu (2004) for a presentation in an instrumental linear regression with $i.i.d.$ disturbances and Coudin and Dufour (2008) for equivalence (with probability one) between both definitions.

For practical use, we also introduce a two-step estimator $\hat{\theta}_n^{2S}(\Omega_n)$ as any solution of the problem

$$\min_{\theta \in \mathbb{R}^p} s(f(y, X, \theta))' Z_\theta \Omega_n \big(s(f(Y, X, \hat{\theta}_n)), Z_{\hat{\theta}_n}\big) Z_\theta' s(f(y, X, \theta)), \tag{7.30}$$

where $\hat{\theta}_n$ is a first stage consistent estimator.

In the following, we show that the IV sign-based estimators defined in equations (7.29) and (7.30) are consistent and asymptotically normal if the parameter is identified.

7.2. Consistency

We first prove the consistency of IV sign-based estimators when the auxiliary regressors are integrable and continuous functions of the parameter θ and of some l-vector process $v_t,\ t = 1, 2, \ldots,$ on which the mixing conditions are imposed. Let $h_t\ :\ \Theta \times \mathbb{R}^l \to \mathbb{R}^k,\ \forall t,$

$$z_t(\theta) = h_t(\theta, v_t),\ t = 1, \ldots. \tag{7.31}$$

We assume that the following conditions hold.

Assumption A7 MIXING. *$\{W_t^v = (y_t, x_t', v_t')\}_{t=1,2,\ldots}$ is α-mixing of size $-r/(r-1)$ with $r > 1$.*

Assumption A8 CONTINUITY OF F. *$f(y_t, x_t, \theta)$ is measurable, a.e. continuous in θ with $P[f(y_t, x_t, \theta) = 0] = 0,\ \forall \theta \in \Theta$.*

Assumption A9 BOUNDEDNESS AND CONTINUITY.

a) $z_t(\theta) = (z_{1t}(\theta), \ldots, z_{pt}(\theta))'$ *and* $E|z_{ht}(\theta)|^{r+1} < \Delta < \infty,\ h = 1, \ldots, k,\ t = 1, \ldots, n,\ \forall n \in \mathbb{N},\ \forall \theta \in \Theta.$

b) $z_{ht}(\theta)$ *is a.e. continuous in* $\theta, \forall t.$

c) $P[z_{ht}(\theta) = 0] = 0,\ \forall \theta \in \Theta, \forall t.$

Assumption A10 COMPACTNESS. $\theta \in$ *Int*(Θ)*, where* Θ *is a compact subset of* $\mathbb{R}^p$*.*

Assumption A11 POINT IDENTIFICATION.

$$\lim_{n\to\infty} E\left[\frac{1}{n}\sum_t s\big(f(y_t, x_t, \theta)\big) \otimes z_t(\theta)\right] = 0 \Rightarrow \theta = \theta_0$$

Assumption A12 UNIFORMLY POSITIVE DEFINITE WEIGHT MATRIX. $\Omega_n(\theta)$ *is symmetric positive definite for all* θ *in* Θ*.*

Assumption A13 LOCALLY POSITIVE DEFINITE WEIGHT MATRIX. $\Omega_n(\theta)$ *is symmetric positive definite for all* θ *in a neighborhood of* θ_0*.*

The mixing condition (Assumption A7) is imposed on a underlying process, $\{v_t\}_{t=1,2,\ldots,}$ because the instruments are functions of the parameter. Assumptions A8 and A9 contain the regularity conditions required on the functions f and h_t. Remark in particular that the sets of zeros are assumed to be negligible. Assumption A10 is the classical compactness condition. Assumptions A11, A12 and A13 are classical and required for identification. Then we have the following property.

Theorem 7.1 CONSISTENCY. *Under model* (2.1) *with the Assumptions A3 and A7-A12, any IV sign-based estimator defined by* (7.29) *is consistent.*

When Assumption A12 is replaced by Assumption A13, the two-step estimators defined in (7.30) are consistent. Consistency is established without requiring second-order moment existence of the disturbances u_t. Indeed, the disturbances appear in the objective function only through their sign transforms which possess finite moments at any order. Consequently no additional restriction should be imposed on the disturbance process. Those points also entail a more general CLT than usual.

7.3. Asymptotic normality

Asymptotic normality requires some additional assumptions.

Assumption A14 UNIFORMLY BOUNDED DENSITIES. $\exists g_U < +\infty$ *such that* $, \forall n \in \mathbb{N}, \forall \lambda \in \mathbb{R},$

$$\sup_{1\le t\le n} |g_t(\lambda|x_1, \ldots, x_n)| < g_U,\ a.s.$$

Assumption A15 DIFFERENTIABILITY OF f. *f is a.e. continuously differentiable in θ and* $E||\frac{\partial f}{\partial \theta'}|_\theta|| < +\infty,\ \forall \theta \in \Theta$.

Assumption A16 MIXING WITH $r > 2$. *$\{W_t = (y_t, x_t', v_t')\}_{t=1,2,\ldots,}$ is α-mixing of size $-r/(r-2)$ with $r > 2$.*

Assumption A17 DIFFERENTIABILITY OF h. *$z_t = h_t(\theta, v_t)$ and h_t is a.e. continuously differentiable in θ and $E||\frac{\partial h_t}{\partial \theta'}|_\theta|| < +\infty,\ \forall \theta \in \Theta,\ \forall t = 1, \ldots, n,\ \forall n \in \mathbb{N}$.*

Assumption A18 DEFINITE POSITIVENESS. *$J_n(\theta_0)$ is $k \times k$ and uniformly positive definite in n and converges to a definite positive symmetric matrix J, where, $J_n(\theta) =$ var $\left[\frac{1}{\sqrt{n}}\sum_t^n s(u_t)h_t(\theta, v_t)\right]$.*

Assumption A19 DEFINITION OF L_n. *$L_n(\theta_0)$ is a $p \times k$ matrix defined as:*

$$
\begin{aligned}
L_n(\theta) &= \frac{1}{n}\sum_t E\left[h_t(\theta, v_t)\frac{\partial f(y_t, x_t, \theta)}{\partial \theta'} g_t\big(f(y_t, x_t, \theta)|z_1, \ldots, z_n\big)\right] \\
&\quad + \frac{1}{n}\sum_t E\left[\frac{\partial h_t}{\partial \theta'} s\big(f(y_t, x_t, \theta)\big)\right].
\end{aligned}
$$

$L_n'(\theta_0)\Omega_n L_n(\theta_0)$ is nonsingular uniformly in n.

Assumption A16 is the classical mixing condition required in asymptotic normality proofs. Assumptions A15, A17 and A19 are regularity conditions for nonlinear setups. Assumption A14 is usual in the LAD and quantile theory: bounded variance conditions (horizontal spread) are replaced by bounded vertical spreads. Assumption A18 is classical. We see in Assumption A19 that $L_n(\theta)$ has a second term induced by the fact that the instruments depend on the parameter. Then, we have the following theorem.

Theorem 7.2 ASYMPTOTIC NORMALITY. *Under the conditions for consistency and Assumptions A14-A19 we have:*

$$S_n^{-1/2}\sqrt{n}\big(\hat{\theta}_n(\Omega_n) - \theta_0\big) \xrightarrow{d} N(0, I_p) \tag{7.32}$$

where

$$S_n = \ [L_n(\theta_0)\Omega_n L_n(\theta_0)']^{-1} L_n(\theta_0)\Omega_n J_n \Omega_n L_n(\theta_0)'[L_n(\theta_0)\Omega_n L_n(\theta_0)']^{-1}.$$

When $\Omega_n = \hat{J}_n^{-1}$,

$$[L_n(\theta_0)\hat{J}_n^{-1} L_n(\theta_0)]^{-1/2}\sqrt{n}\big(\hat{\theta}_n(\hat{J}_n^{-1}) - \theta_0\big) \xrightarrow{d} N(0, I_p). \tag{7.33}$$

Theorem **7.2** holds in particular for classical instrumental setups when the instruments Z do not depend on θ. In such a case, $L_n(\theta)$ simplifies to

$$L_n(\theta) = \frac{1}{n}\sum_t E\left[z_t \frac{\partial f(y_t, x_t, \theta)}{\partial \theta'} g_t(f(y_t, x_t, \theta)|z_1, \ldots, z_n)\right]. \tag{7.34}$$

This result extends the classical sign-based estimator asymptotic normality established in Coudin and Dufour (2008) for nonlinear and instrumental regressions. Note again the existence of the second-order moment disturbances is not required. The sign asymptotic normality holds for heavy-tail distributions whereas usual estimators, such as the 2SLS estimator, do not. The dispersion measure adapted to sign-based estimators do not refer to the error variance but to the (inverse of the) error density evaluated at zero. This alternative dispersion measure, called the "diffusivity", is involved in Cramér-Rao type lower bound for median-unbiased estimators; see Coudin and Dufour (2008), Sung, Stangenhaus, and David (1990) and So (1994).

The properties of consistency and asymptotic normality entirely rely on the identification assumption whereas the sign-based inference presented previously does not. This provides the occasion to recall the main message of the weak IV literature: when some identification failure or the presence of weak instruments are suspected, tests based on the asymptotic behavior of estimators should be avoided. Inference should be based on test statistics that are robust to identification failure such as IV sign-based statistics. The next section illustrates by a simulation study, how important it can be to use the exact distribution of such robust statistics.

8. Simulation study

In this section, we present simulation studies comparing the performance of sign-based methods with usual instrument-based techniques. We consider the basic sign-based statistic $D_S(\theta, Z, (Z'Z)^{-1})$ (denoted BS) and a split-sample based one that aims to overcome possibly power loss when "many instruments" are used (SSS). We compare tests based on those two statistics with Wald tests based on the 2SLS estimator and the 2SLAD estimator (both estimators are unreliable in the presence of weak instruments), and with some tests that are "robust to weak instruments". Those robust tests rely on the Anderson-Rubin statistic (AR) [Anderson and Rubin (1949)], the Anderson-Rubin statistic with split-sample (SSAR) [Dufour and Jasiak (2001)], the score statistic proposed by Kleibergen (2002) (K) and the score statistic corrected for heteroskedasticity (KLM) [Kleibergen (2007)]. We use the following linear model taken from Kleibergen (2002) with different numbers of instruments, degrees of identification and various disturbance behaviors:

$$y = Y\theta + \epsilon$$
$$Y = X\mathit{\Pi} + V,$$

where n is the number of observations, $y,\ Y\ :\ n\times 1,\ X\ :\ n\times k,\ X\sim\mathcal{N}(0,I_k\otimes I_n)$, $\Pi : k\times 1, \theta=0$. In $\Pi=(\pi_1,\ldots,\pi_k)'$, four different values of π_1 are considered: 1 (strong valid instrument), 0.5 (instrument of mild strength), 0.1 (weak instrument), and 0 (no identification). Other components of Π are set to zero. The number of instruments k alternatively equals 1, 5 or 10 in view of studying the effect of including irrelevant instruments.

We wonder what the test performances are for various schemes of disturbances. Therefore, we do not restrict on $i.i.d.$ normal disturbances. We also study heavy-tailed disturbances and heteroskedastic schemes. We use the four following data generating processes:

Case 1: $i.i.d.$ normal disturbances:
$(\epsilon, V)\sim\mathcal{N}(0,\Sigma\otimes I_n)$, $\Sigma=\begin{pmatrix}1 & .99\\ .99 & 1\end{pmatrix}$.

Case 2: $i.i.d.$ Cauchy disturbances:
$(\epsilon^1, V^1)\sim\mathcal{C}$ and $(\epsilon_t, V_t)'=\Sigma(\epsilon_t^1, V_t^1)'$, with $\Sigma=\begin{pmatrix}1 & .99\\ .99 & 1\end{pmatrix}$.

Case 3: some instruments affect the shape of the structural error ϵ :
$(\epsilon^1, V)\sim\mathcal{N}(0,\Sigma\otimes I_n)$, $\epsilon_t=x_{t1}^2\epsilon_t^1, t=1,\ldots,T$.

Case 4: the endogenous variable affects the shape of ϵ:
$(\epsilon^1, V)\sim\mathcal{N}(0,\Sigma\otimes I_n)$, $\epsilon_t=Y_t^2\epsilon_t^1, t=1,\ldots,T$.

Cases 1 and 2 illustrate the effect of a departure from normality on the different tests: homoskedastic disturbances, which are normally distributed in case 1 and Cauchy distributed in case 2. In normal cases, with one instrument, the K statistic which equals the AR is optimal. We wonder what happens when normality is relaxed and especially when the disturbances possess heavy tails. The next DGPs (cases 3 and 4) illustrate heteroskedasticity. In case 3, the instruments affect the variance of the structural error. In case 4, the endogenous variable affects the variance of the structural error. We illustrate how the classical tests (K, AR) fail in the presence of heteroskedasticity and we focus on comparing sign-based tests to the KLM tests that are corrected for heteroscedasticity. Remark that for the four cases, the mediangale Assumption A1 holds and sign-based methods do exactly control levels for any sample size.

8.1. Size

We first investigate level distortions. We consider the testing problem: $H_0:\theta_0=0$ versus $H_1:\theta_0\neq 0$, and report empirical rejection frequencies for tests of level .05. Empirical sizes are computed using 10000 simulations. Bootstrap and Monte Carlo methods are both based on 2999 replicates. For split-sample statistics (SSAR and SSS), 15 observations are used for the first stage and 35 for the second stage.

Sign-based tests (BS, SSS) are the only ones that have perfectly controlled levels in the four presented cases. Empirical sizes of sign-based tests equal the nominal size. In contrast, empirical sizes of Wald tests (2SLAD, 2SLS) greatly suffer from the small number of observations, the weakness of the instruments and the presence of irrelevant instruments. The empirical sizes of the AR, SSAR and K tests are smaller than the Wald-type test ones in homoskedastic setups because their asymptotic levels equal the nominal one whatever the strength and the number of instruments. However, they are affected by finite-sample distortions and loose their relevance in heteroskedastic setups. Finally, tests based on the KLM statistic involving a White-type correction for heteroskedasticity have empirical sizes close to the nominal one for setup 3, but this is no longer true when endogeneity affects the variance of the structural error (setup 4).

Simulations confirm the theory. Sign-based tests allow to control test levels for a very wide range of setups and for any sample size. They are the only ones that are robust to heteroskedasticity of unknown form.

Table 1. Empirical sizes: n=50.

Case 1 : $i.i.d.$ normal distribution												
nb inst.	k=1				k=5				k=10			
π_1	1	.5	.1	0	1	.5	.1	0	1	.5	.1	0
W2SLS	.087	.123	.375	.911	.315	.708	.994	1.00	.548	.939	1.00	1.00
W2SLAD	.028	.019	.001	.000	.161	.352	.691	.715	.296	.595	.873	.889
AR	.059	.059	.059	.059	.067	.067	.067	.067	.088	.088	.088	.088
SSAR	.116	.116	.116	.116	.095	.096	.097	.097	.085	.086	.084	.084
K	.059	.059	.059	.059	.057	.057	.056	.070	.060	.060	.060	.088
KLM	.048	.048	.048	.048	.024	.024	.024	.036	.016	.016	.016	.032
BS	.050	.050	.050	.050	.045	.045	.045	.045	.056	.056	.056	.056
SSS	.052	.052	.052	.052	.049	.048	.047	.047	.052	.050	.051	.051
Case 2 : $i.i.d.$ Cauchy distribution												
π_1	1	.5	.1	0	1	.5	.1	0	1	.5	.1	0
W2SLS	.477	.607	.822	.937	.987	.998	1.00	1.00	1.00	1.00	1.00	1.00
W2SLAD	.001	.001	.000	.000	.037	.037	.038	.036	.045	.047	.048	.047
AR	.061	.061	.061	.061	.063	.063	.063	.063	.081	.081	.081	.081
SSAR	.121	.121	.121	.121	.103	.103	.102	.102	.080	.082	.081	.081
K	.061	.061	.061	.061	.054	.054	.055	.066	.066	.067	.067	.077
KLM	.019	.019	.019	.019	.034	.034	.034	.032	.027	.028	.028	.029
BS	.051	.051	.051	.051	.053	.053	.053	.053	.056	.056	.056	.056
SSS	.050	.050	.050	.050	.047	.047	.047	.047	.056	.053	.056	.055
Case 3 : instruments affect the shape of error distribution												
π_1	1	.5	.1	0	1	.5	.1	0	1	.5	.1	0
W2SLS	.101	.129	.203	.213	.140	.256	.475	.493	.160	.328	.674	.700
W2SLAD	.021	.015	.004	.003	.048	.039	.017	.016	.088	.081	.047	.044
AR	.417	.417	.417	.417	.249	.249	.249	.249	.223	.223	.223	.223
SSAR	.510	.510	.510	.510	.280	.215	.184	.179	.179	.131	.111	.111
K	.417	.417	.417	.417	.329	.263	.159	.153	.357	.259	.129	.120
KLM	.029	.029	.029	.029	.026	.034	.040	.040	.032	.038	.043	.042
BS	.053	.053	.053	.053	.048	.048	.048	.048	.057	.057	.057	.057
SSS	.053	.053	.053	.053	.055	.051	.052	.050	.051	.051	.053	.054
Case 4 : endogeneity affects the shape of error distribution												
π_1	1	.5	.1	0	1	.5	.1	0	1	.5	.1	0
W2SLS	.744	.519	.234	.216	.898	.849	.821	.822	.923	.967	.972	.972
W2SLAD	.012	.006	.001	.001	.030	.028	.027	.026	.056	.059	.062	.064
AR	.526	.220	.068	.061	.300	.128	.072	.069	.323	.162	.084	.080
SSAR	.527	.269	.128	.121	.282	.135	.097	.096	.221	.108	.081	.079
K	.526	.220	.068	.061	.406	.128	.068	.068	.497	.169	.081	.082
KLM	.321	.126	.032	.028	.207	.077	.040	.039	.055	.068	.044	.041
BS	.051	.051	.051	.051	.044	.044	.044	.044	.054	.054	.054	.054
SSS	.050	.050	.050	.050	.049	.052	.051	.051	.049	.051	.050	.050

8.2. Power

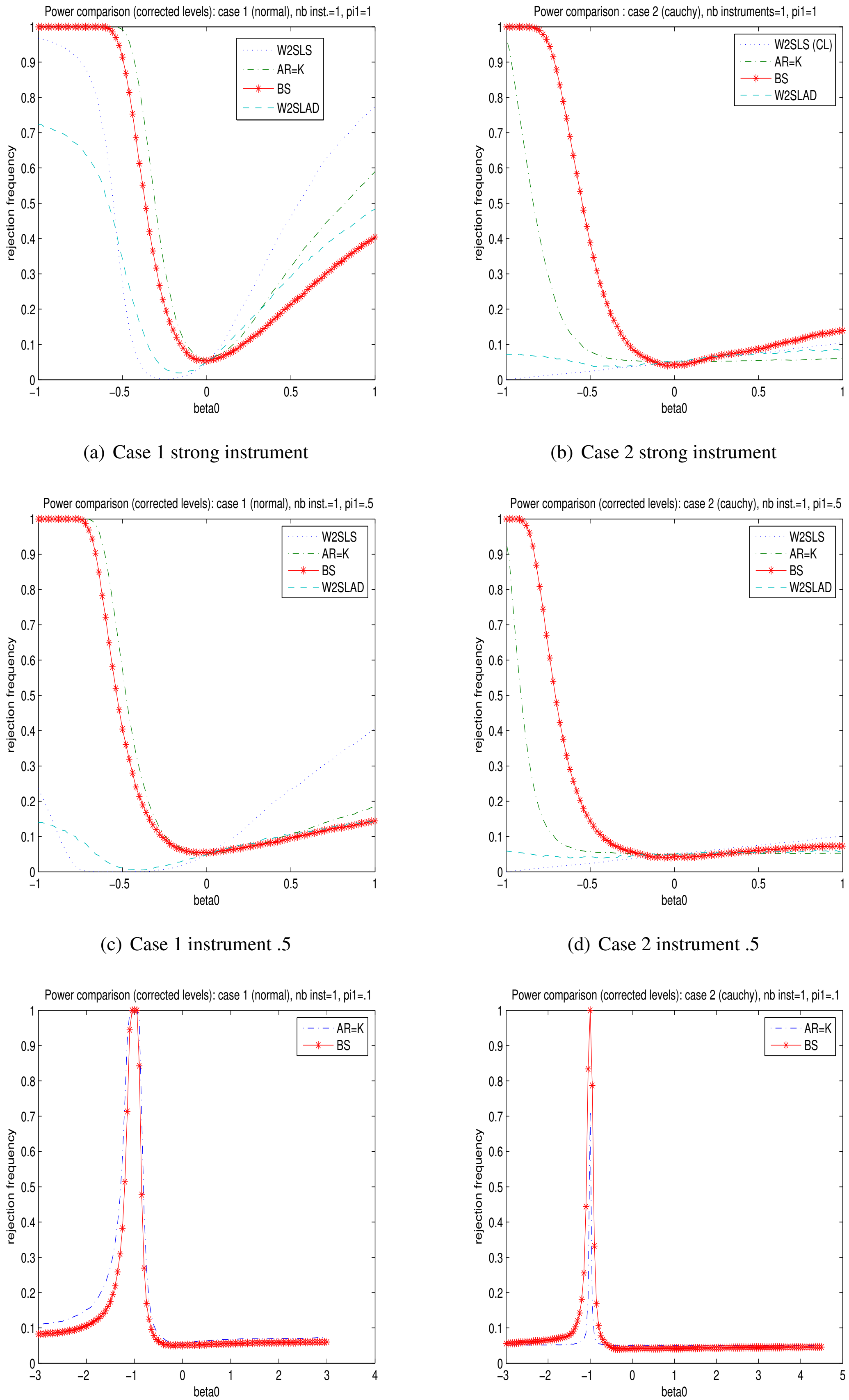
Then, we compare the power of these tests. Tests of $H_0 : \theta = 0$ are performed on data obtained by letting vary θ. Simulated power is given by a graph with θ in abscissa; see Figures 1, 2, 3, 4. The power functions presented here are locally adjusted for the level when needed, which allows comparisons between methods. However, we should keep in mind that only sign-based tests do exactly control the level for any sample size. All results concerning homoskedastic or heteroskedastic setups with a given number of instruments and for various instrument strength are contained in a single figure. In Figures 1 and 2, errors are homoskedastic, either normal (first column), either Cauchy (second column). The number of instruments equals one for Figure 1, and five for Figure 2. Therefore, comparing both columns illustrates which effect a departure from normality (here Cauchy disturbances) entails on the test powers. The effect of heteroskedasticity is then illustrated by Figures 3 (model with one instrument) and 4 (model with five instruments). We are particularly interested in comparing the sign-based method to the KLM method (and 2SLAD, 2SLS for strong instruments) which is corrected for heteroskedasticity since the K and the AR methods are not.

Let us now examine the results. In a model with one instrument (Figure 1), the K statistic and the AR statistic are equal. The AR statistic is best for the $i.i.d.$ normal case 1 but the sign-based power curve is not far from that optimal power curve (first column of Figure 1). With Cauchy distributions (case 2, column 2 in Figure1), the sign-based power curve is far above all the others. This holds regardless of instrument strength. The power curves of Wald tests based on the 2SLS and the 2SLAD estimators are also reported when the instruments are strong. In case 1, these methods are biased; in case 2, they do not present power anymore.

The AR procedure and the sign-based procedure loose power as the number of (irrelevant) instruments included in the model increases. Figure 2 illustrates the power curves when the model involves five instruments. For the $i.i.d.$ normal case (case 1, column 1 in Figure2), the K statistic, which now differs from the AR statistic, does not encounter this loss of power and leads to the highest power curve whereas both the sign-based power curve and the AR-based one stand lower. However, as soon as we turn to the Cauchy setup (case 2, column 2 in Figure2), the sign-based statistic yields again the highest power. This holds regardless of instrument strength. The two methods involving a split-sample (SSAR and SSS) do not present good results because of the limited number of observations. Here, the sample size is 50. First step regressions involve only 15 observations and second step regressions 35 observations. However, the corresponding power curves generally follow the same tendencies as the power curves of the corresponding statistic *without* split-sample.

Results are very clear in Figures 3 and 4 (heteroskedastic setups: case 3 and 4). Sign-based methods exhibit there more power than all the other studied methods which are robust to weak instruments (AR, K) included methods corrected for heteroskedasticity (KLM). In the presence of strong instruments, Wald tests based on 2SLAD and 2SLS have higher power than sign-based methods. However, the Wald tests are clearly biased and they are no longer valid as soon as the strength of the instruments decreases.

In conclusion, sign-based tests present good power properties for a wide range of processes. They are not far from the optimal AR test in $i.i.d.$ normal case and they provide more power than other studied methods in setups involving heavy-tailed distributions, heteroskedasticity or nonlinear dependence. They still provide power under some general endogeneity schemes, especially when the endogeneity affects the shape of the structural error distribution without affecting its sign.



(a) Case 1 strong instrument

(b) Case 2 strong instrument

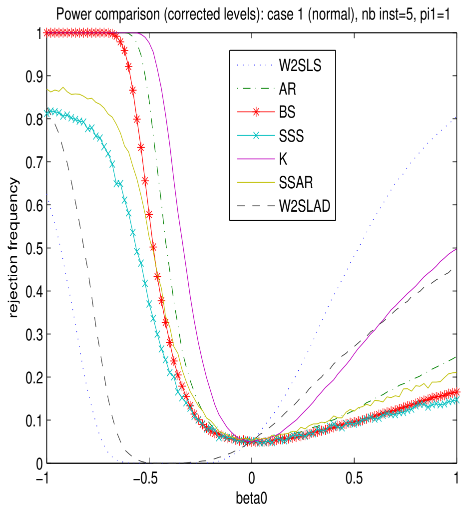
(c) Case 1 instrument .5

(d) Case 2 instrument .5

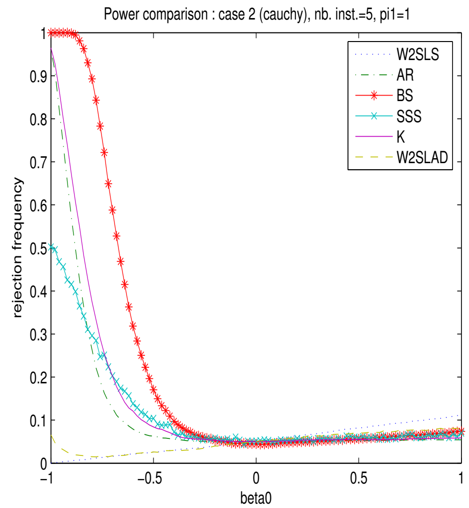
(e) Case 1 weak instrument

(f) Case 2 weak instrument

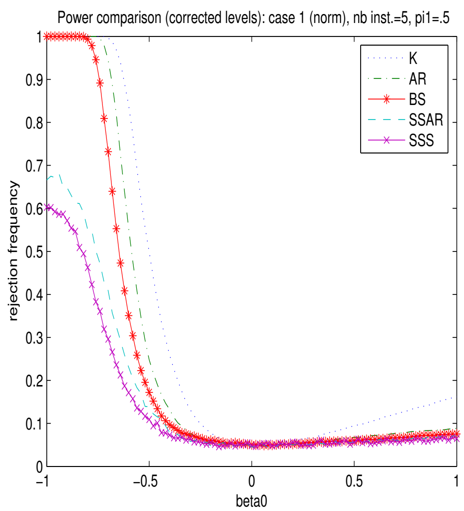
Figure 1. Power functions: model with one instrument.



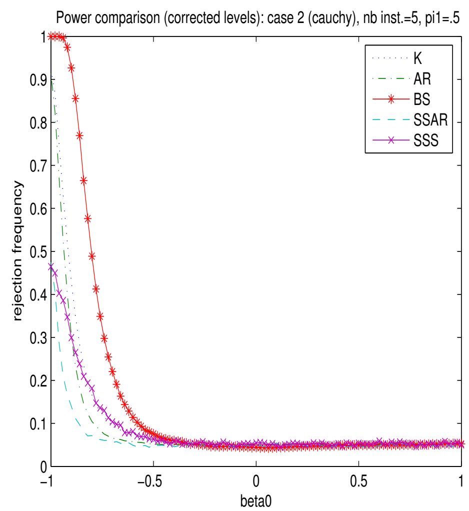
(a) Case 1 strong instrument



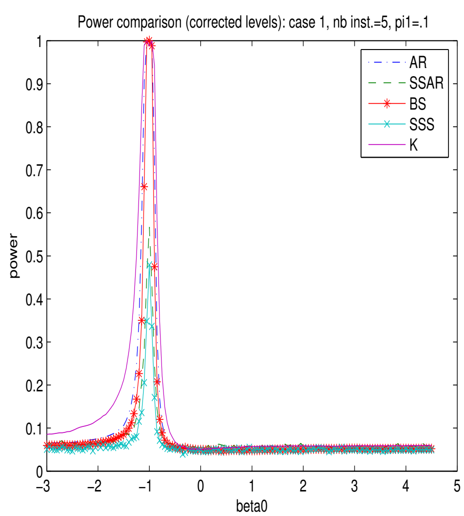
(b) Case 2 strong instrument



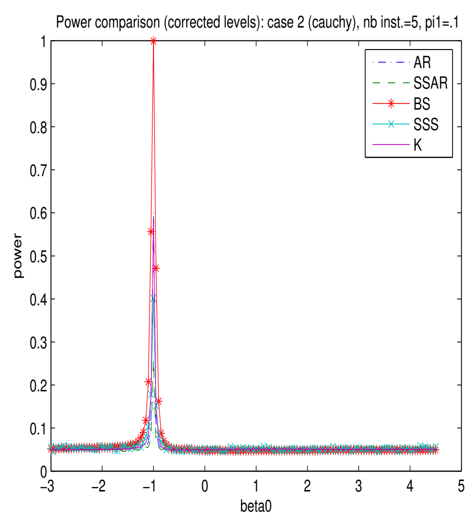
(c) Case 1 mild instrument



(d) Case 2 mild instrument

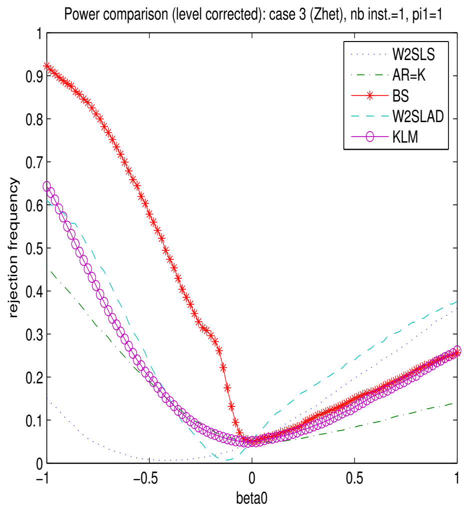


(e) Case 1 weak instrument

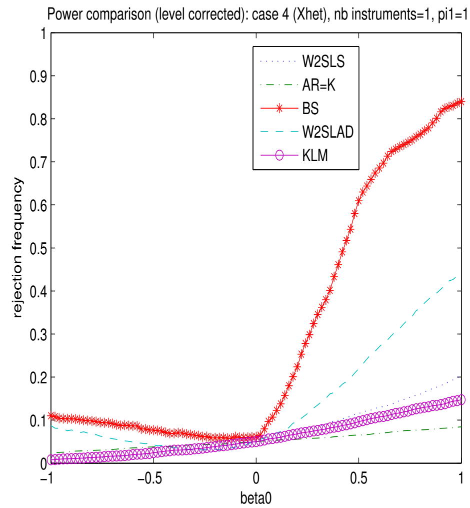


(f) Case 2 weak instrument

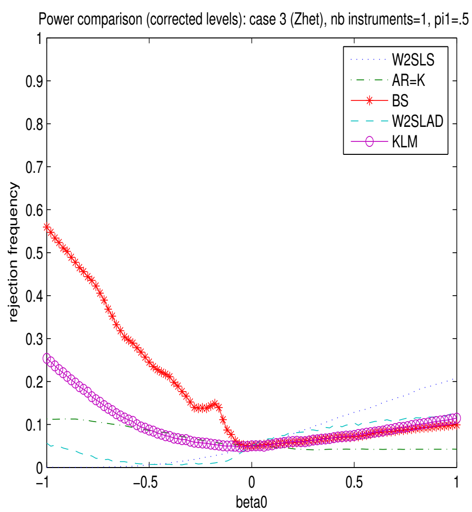
Figure 2. Power functions: model with 5 instruments.



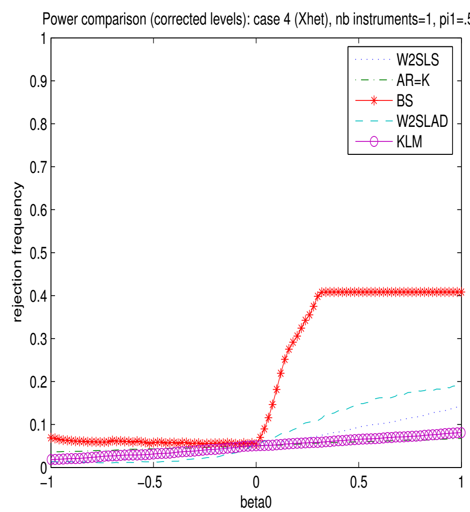
(a) Case 3 strong instrument



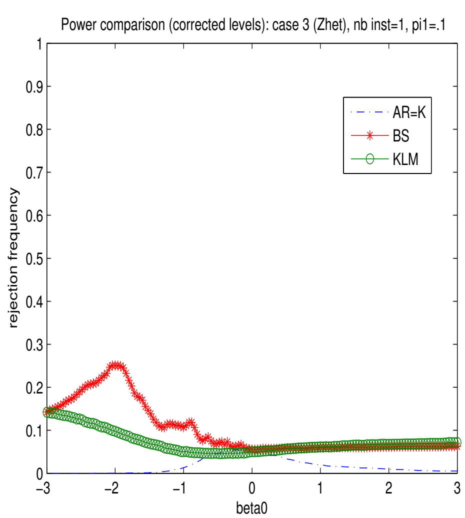
(b) Case 4 strong instrument



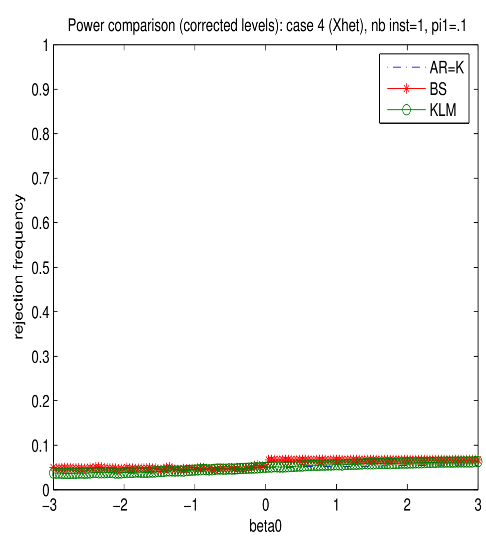
(c) Case 3 instrument .5



(d) Case 4 instrument .5

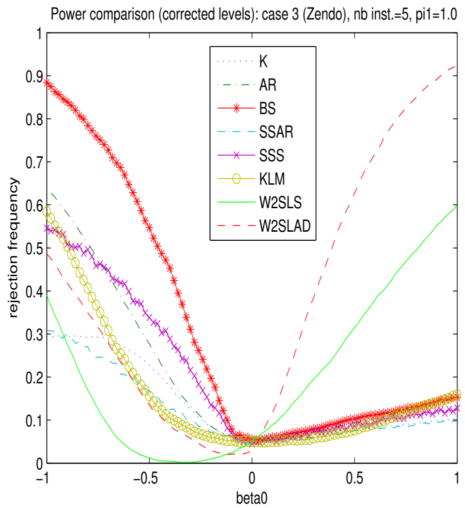


(e) Case 3 weak instrument

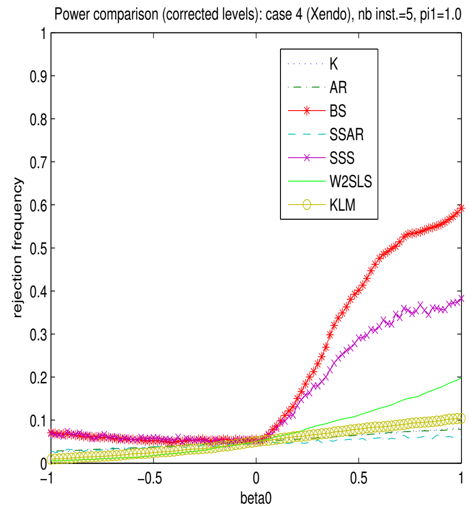


(f) Case 4 weak instrument

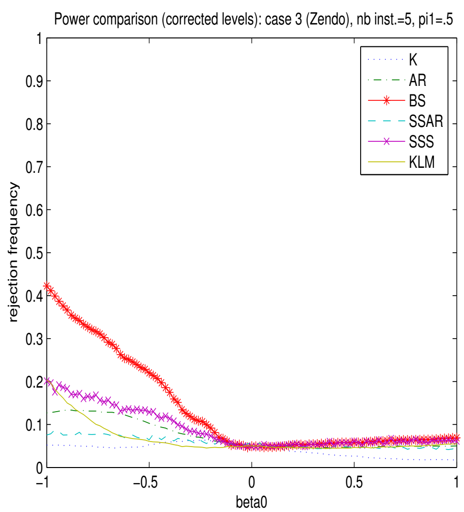
Figure 3. Power functions: model with one instrument: heteroscedastic cases.



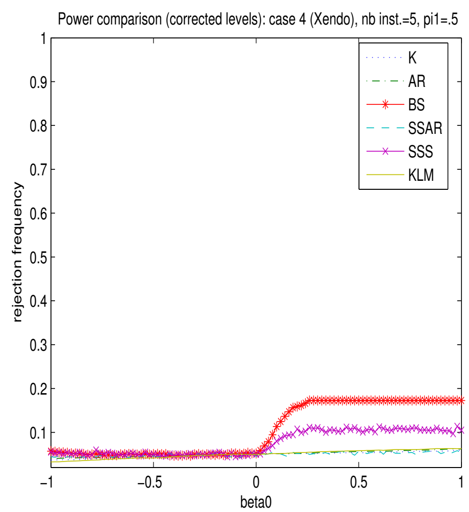
(a) Case 3 strong instrument



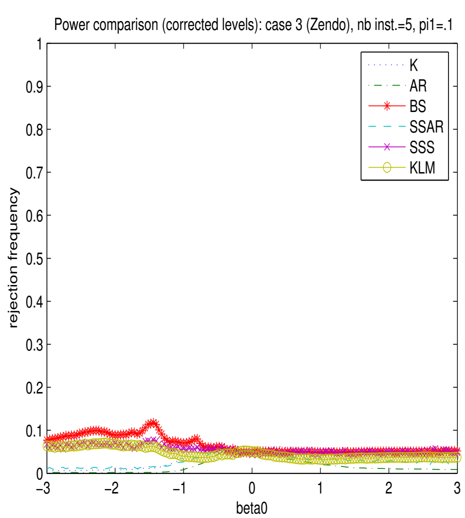
(b) Case 4 strong instrument



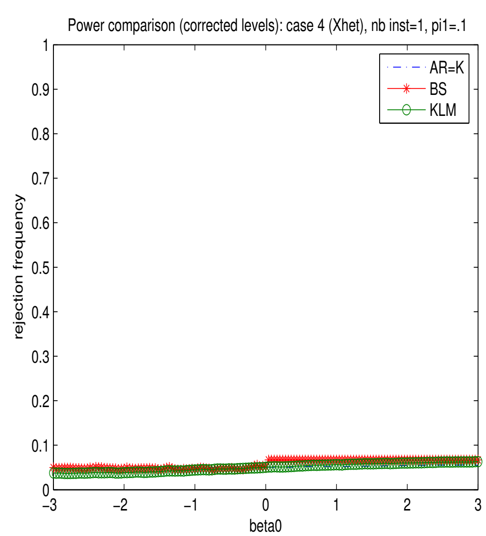
(c) Case 3 instrument .5



(d) Case 4 instrument .5



(e) Case 3 weak instrument



(f) Case 4 weak instrument

Figure 4. Power functions: model with five instruments: heteroscedastic cases.

9. Application: schooling returns

In this section, we apply the method proposed above to study the effect of education on earnings [Angrist and Krueger (1991), Angrist and Krueger (1995), Bound, Jaeger, and Baker (1995), Staiger and Stock (1997), Dufour and Jasiak (2001), Kleibergen (2002, 2005, 2007), etc.].[14] Angrist and Krueger (1991) consider an earning equation where the log weekly wage (y_t) is explained by the year number of schooling (x_t) and other covariates (such as the year of birth, age, age squared, race, metropolitan statistical area...). They propose several specifications depending on the included covariates. Further, they use the interactions between the quarter of birth and the year of birth as instruments for correcting the education endogeneity. However, the relation between the instruments and the endogenous variable is apparently weak.

We restrict here on the Angrist and Krueger (1991)'s model specification with dummies for the year of birth as explanatory variables. The data set comes from the 1980 census 5% public-use sample and is composed of $n = 329500$ men born 1930-39.

$$y_i = \beta_0 x_i + \sum_{k=1}^{10} \beta_k d_{ki} + \epsilon_i,\ i = 1, \dots, n, \tag{9.1}$$

where d_k are dummies for the year of birth. Further, the 30 interactions between the quarter and the year of birth constitute the available "excluded" instruments to correct for the schooling endogeneity. F-statistic for instrument relevance equals 1.573 (with asymptotic p-val= .024), which is low enough to suspect the presence of weak instruments.

We apply split-sample sign-based inference method and compute valid confidence intervals for the education parameter. More precisely, the sample is divided into two parts (1) and (2). With the first part of the sample, we choose the form of quasi-optimal instruments: the year number of schooling is regressed on instruments by OLS. With the second subsample, we construct sign-based statistic using a fitted education. The split-sample sign-based statistics rely on the 11 following moment equations:

$$E[s(y_i^{(2)} - \beta_0 x_i^{(2)} - \sum_{k=1}^{10} \beta_k d_{ki}) \times \tilde{z}_{ji}^{(2)}] = 0, \quad \text{for } i = 1, \dots, n_2,\ j = 1, \dots, 11; \tag{9.2}$$

where $\tilde{z}_{ji}^{(2)} = d_{ji},\ j = 1, \dots, 10$ and $\tilde{z}_{11i}^{(2)}$ is the fitted education. We follow Dufour and Jasiak (2001), and use 10% of the sample for the first stage and 90% for the second one. Two split-sample sign-based statistics are considered. The first one combines moment equations in a classical quadratic GMM form (SSS90). In the second one (TSS90), moment equations are combined following Tippett (1931). Then, Bonferroni-type induced tests are performed using $\alpha_m = \alpha/p$. The idea behind is that quadratic combination of orthogonality conditions refers to some local optimal-

[14] Other questions raised by these data include, for example, the impossibility of a punctual nonparametric identification with discrete instruments [Chesher (2005)] and the problem of many instruments [Hansen, Hausman, and Newey (2005)].

Table 2. Confidence intervals for schooling returns.

CI	95%	90%	80%
Wald OLS	[.070, .072]	[.071, .072]	[.071, .071]
Wald 2SLS	[.058, .120]	[.063, .115]	[.069, .110]
Wald 2SLAD*	[-.002, .079]	[.004, .073]	[.012, .065]
AR	[.014, .180]	[.022, .169]	[.033, .157]
K	[.054, .133]	[.060, .126]	[.068, .119]
TSS90	[.034, .045]	[.036, .044]	[.037, .043]
SSS90	[.035, .045]	[.036, .041]	[.038, .039]

* W2SLAD CI are obtained by design matrix bootstrap, with 99 replicates [Buchinsky (1998)].

ity around the true value of the parameter. In a badly identified setup such as here, other type of combinations like the Tippett's one, may provide better overall properties and smaller confidence intervals.

Table 2 contains $95\%-$confidence intervals obtained with SSS90 and TSS90 but also the Anderson and Rubin statistic (AR), Kleibergen score statistic (K) and Wald (non reliable) CI based on the OLS and the 2SLS estimators. We also report in Table 3 OLS, 2SLS, LIML, SSIV and sign estimates for the return to education.[15]

Projection sign-based confidence intervals obtained using the SSS90 and the TSS90 statistics have smaller spreads than the asymptotic ones based on the AR and K statistics and they are theoretically valid. Moreover, they tend to accept smaller values of the return to education. Table 3 on estimates confirms that point. Sign-based estimates that are very close to 2SLAD estimates, suggest a return to schooling around 4% which is smaller than usually admitted. Such a figure is in adequation with a positive ability bias as expected by the theory.

Then we redo the same experiment on subsamples of 10000 and 2000 observations drawn from the initial sample. We wonder what happens when the sample size gets smaller. Confidence intervals results are reported in Table 4 and estimates in Table 5. We only consider procedures that are robust to weak instruments: K, AR, SSS90 (with 999 replicates) and TSS90 (with 879 replicates).

[15]The CI are smaller than those found by Chernozhukov, Hansen, and Jansson (2009) who exploited a GMM statistic based on the 40 moment equations and included in their model more explanatory variables. We use simulated annealing with different starting points. They used a MCMC algorithm with different starting points.

Table 3. Estimates for schooling returns.

	OLS	2SLS	LAD	2SLAD
β_0	.071	.089	.066	.039
	LIML	SSIV90	SSS90	
β_0	.093	.018	.039	

Table 4. Confidence intervals for schooling returns: subsamples n=10000 and n=2000.

CI	95%	90%	80%
		n=10000	
K	[-1,1]	[-1,.222]∪[.239,1]	[-1, -.300]∪[-.012,.145]∪[.404,1]*
AR	[-1,1]	[-.636,.664]	[-.291,.395]
TSS90	[-.190,.109]	[-.110,.083]	[-.034,.049]
SSS90	[-1,1]	[-1,1]	[-1,.236]
		n=2000	
K	[-1,1]	[-1,.073]∪[.106,1]	[-.563, .016]∪[.160,.541]*
AR	[-1,1]	[-1,1]	[-1,.154]∪[.562,1]
TSS90	[-.392,.135]	[-.216,.075]	[-.130,.043]
SSS90	[-1,1]	[-1,1]	[-1,1]

* CIs can be reduced by combining with a J test [Kleibergen (2007)].

Table 5. Estimates for schooling returns: subsamples n=10000 and n=2000.

			n=10000				
β_0	OLS	2SLS	LAD	2SLAD	LIML	SSIV90	SSS90
	.072	.076	.065	.022	.067	-.012	.022
			n=2000				
β_0	OLS	2SLS	LAD	2SLAD	LIML	SSIV90	SSS90
	.071	.014	.067	.022	-.119	-.013	.023

The Mincer equation (9.1) sets that the education coefficient has an elasticity form. Consequently, this parameter is constrained in the programs to rely between -1 and 1. Then, a confidence interval of $[-1, 1]$ may refer to an (unconditional) "unbounded" confidence interval. Such a confidence interval indicates a badly identified setup and is in accord with the fact that valid confidence intervals have positive probability to be unbounded in nonidentified setups [Dufour (1997)].

The CI spread based on SSS90 and AR statistics increases as the number of observations decreases. 90%-CI based on the AR statistic is bounded for $n = 10000$ whereas for $n = 2000$, the 90%-CI is $[-1, 1]$. The same occurs with 95%-CI based on the SSS90 statistic. The behavior of the K statistic is less clear. As it is a quadratic form of the score of the concentrated log-likelihood, it basically contains information on a slope. Its use is locally justified around the LIML estimator but may follow a somewhat odd behavior outside that neighborhood. The Tippett-sign-based statistic provides the smaller CIs for both subsamples, which indicates that quadratic combinations of orthogonality conditions are not optimal in small subsamples.

Concerning estimates (Table 5), our findings are similar to the whole sample ones. Sign-based estimates are very close to 2SLAD estimates and suggest returns to schooling around 2% in both subsamples which is in adequation with the theoretically expected ability bias.

10. Conclusion

In this paper, we presented a finite-sample sign-based inference system for the parameter of a structural possibly nonlinear model. We introduced a condition of instrument validity with respect to the signs of the structural error. We showed that, under the instruments validity, the distribution of the structural error sign vector is known and does not depend on any nuisance parameter. This allowed us to conduct a Monte Carlo-based inference using on the exact distribution of IV sign-based statistics. The derived joint tests are exact for any sample size and are robust to identification failures. Tests of more general hypothesis and confidence sets are then constructed using projection techniques. Our approach is in the spirit of Anderson and Rubin (1949). The IV sign-based statistics we studied can be constructed from auxiliary regressions of the constrained signs on auxiliary instruments. We also considered the problem of approaching the optimal set of instruments to include in the model in case of overidentification using two different optimality concepts (point and local optimality). Finally, IV sign-based estimators are presented. They turn to be consistent and asymptotically normal when identification holds under weaker assumptions than the ones required in the 2SLAD asymptotic theory. Besides, they can directly be associated with previous sign-based inference, which avoids one to use complicated methods such as the bootstrap. By construction, the level of IV sign-based tests is controlled and simulations indicate that those tests perform better than usual ones (including methods that are robust to weak instruments or identification failures) in finite samples, when the data are heterogenous, heteroskedastic or when endogenous variables affect the structural error distribution without affecting its sign. Finally, sign-based inference is applied to the Angrist and Krueger's returns to schooling problem. Sign-based estimate of the return to schooling is around 4% and projection-based confidence intervals, besides being more robust, are more precise than those based on the AR or the K statistics. In small samples, it seems that Tippett-type combination of orthogonality conditions provides better properties than usual quadratic combination and leads to more precise confidence intervals.

Appendix

A. Proofs

A.1. Proof of Proposition 3.1

Consider the vector $[s(u_1), s(u_2), \ldots, s(u_n)]' \equiv (s_1, s_2, \ldots, s_n)'$. From Assumption A1, we derive the two following equalities:

$$\begin{aligned} P(u_t > 0|Z) &= \mathsf{E}[P(u_t > 0|u_{t-1}, \ldots, u_1, Z))] = 1/2, \\ P(u_t > 0|s_{t-1}, \ldots, s_1, Z) &= P(u_t > 0|u_{t-1}, \ldots, u_1, Z) = 1/2, \forall t = 2, \ldots, n. \end{aligned}$$

Further, the joint density of $(s_1, s_2, \ldots, s_n)'$ can be written:

$$\begin{aligned} l(s_1, s_2, \ldots, s_n|Z) &= \prod_{t=1}^{n} l(s_t|s_{t-1}, \ldots, s_1, Z) \\ &= \prod_{t=1}^{n} P(u_t > 0|u_{t-1}, \ldots, u_1, Z)^{(1-s_t)/2} \\ &\quad \times \{1 - P(u_t > 0|u_{t-1}, \ldots, u_1, Z)\}^{(1+s_t)/2} \\ &= \prod_{t=1}^{n} (1/2)^{(1-s_t)/2}[1 - (1/2)]^{(1+s_t)/2} = (1/2)^n. \end{aligned}$$

Hence, conditional on Z, $s_1, s_2, \ldots, s_n$ are distributed like n i.i.d random variables with distribution:

$$P(s_t = 1) = P(s_t = -1) = \frac{1}{2},\ t = 1, \ldots, n.$$

A.2. Proof of Proposition 4.1

This is a direct application of Neyman-Pearson lemma. The likelihood function of S under H_0 is

$$L_0(s_1, \ldots, s_n) = \prod_{t=1}^{n} p_{0t}^{(1+s_t)/2} p_{0t}^{(1-s_t)/2}$$

and under H_1,

$$L_1(s_1, \ldots, s_n) = \prod_{t=1}^{n} p_{1t}^{(1+s_t)/2} p_{1t}^{(1-s_t)/2}.$$

Hence, after some computations, the loglikelihood ratio becomes

$$\ln\left(\frac{L_1}{L_0}\right) = \sum_{t=1}^{n} (1/2)\left[\ln\left(\frac{p_{1t}(1-p_{1t})}{p_{0t}(1-p_{0t})}\right) + s_t \ln\left(\frac{p_{1t}(1-p_{0t})}{p_{0t}(1-p_{1t})}\right)\right], \tag{A.1}$$

and yields the optimal test for H_0 versus H_1. The most powerful test based on S rejects H_0 when

$$\sum_{t=1}^{n} s_t \ln\left(\frac{p_{1t}(1-p_{0t})}{p_{0t}(1-p_{1t})}\right) > c(\alpha, H_1)$$

where $c(\alpha, H_1) = c - \sum_{t=1}^{T}(1/2)\left[\ln\left(\frac{p_{1t}(1-p_{1t})}{p_{0t}(1-p_{0t})}\right)\right]$ with c derived from Neyman-Pearson condition. $\square$

A.3. Proof of Corollary 4.2

In the regression framework, $(p_{01}, \ldots, p_{0n})$ and $(p_{11}, \ldots, p_{1n})$ are known. As Assumption A1 holds under H_0, we have $p_{0t} = .5$, and under H_1, we can write for $t = 1, \ldots, n.$,

$$p_{1t} = P_{H_1}[f(y_t, x_t, \theta_0) > 0] = P_{H_1}[f(y_t, x_t, \theta_1) > f(y_t, x_t, \theta_1) - f(y_t, x_t, \theta_0)] = 1 - G(h_t),$$

where $h_t = f(y_t, x_t, \theta_0) - f(y_t, x_t, \theta_1)$. Hence, the point-optimal sign-based test of H_0 versus H_1 rejects H_0 when

$$\sum_{t=1}^{n} s(u_t) \ln\left(\frac{1 - G(h_t)}{G(h_t)}\right) > c(\alpha, \theta_1), \tag{A.2}$$

where $(h_1, \ldots, h_n)' = \big(f(y_1, x_1, \theta_1) - f(y_1, x_1, \theta_0), \ldots, f(y_n, x_n, \theta_1) - f(y_n, x_n, \theta_0)\big)'$ and $c(\alpha, \theta_1)$ depending on the level.

A.4. Proof of Proposition 5.1

First, we prove the following lemma.

Lemma A.1 *Let $\{G_n\}_n$ be a sequence of real functions tending uniformly towards G on a compact set $K \subset \mathbb{R}$ and $0 \in int(K)$. Suppose further that G_n and G are differentiable with continuous derivative on K for all n and satisfy $n(G_n(0) - G(0)) \to 0$ and $G_n'(0) - G'(0) \to 0$. Then,*

$$\sup_{y \in B(0, \frac{1}{n})} ||G_n(y) - G(y)|| = o(1/n).$$

Proof of Lemma A.1. Taylor expansions gives

$$G_n(x) = G_n(0) + xG_n'(0) + o(|x|), \forall x \in B(0, 1/n) \cap K, \tag{A.3}$$

and

$$G(x) = G(0) + xG'(0) + o(|x|), \forall x \in B(0, 1/n) \cap K. \tag{A.4}$$

We can write

$$|G_n(x) - G(x)| = |G_n(0) - G(0) + x(G_n'(0) - G'(0)) + o(1/n)|. \tag{A.5}$$

Hence,

$$|G_n(x) - G(x)| \leq |G_n(0) - G(0)| + \frac{1}{n}|G_n'(0) - G'(0)| + o(1/n) \tag{A.6}$$

by majoring $|x|$ by $1/n$. That entails

$$|G_n(x) - G(x)| = o(1/n).\Box \tag{A.7}$$

Let us now consider the problem of testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$. The power function of a sign-based test T conditional on Z is

$$\beta(\theta_1) = P_{\theta_1}\left[T\big(s(f(y, x, \theta_0)), Z\big) > 1 - c_T(Z, \alpha)|Z\right] = P_{\theta_1}[S \in W_\alpha|Z] \tag{A.8}$$

where S is the random variable of the constrained signs and W_α the critical region of the test with level α. In the sequel, we omit to write that all results are conditional on Z. To identify the instruments which maximize the power function in the neighborhood of θ_0, we first derive the sign distribution under H_1. The independence assumption implies that the sign distribution is the product of terms of the form

$$P_{\theta_1}[s_t = s] = P_{\theta_1}[f(y_t, x_t, \theta_0) \geq 0]^{\frac{1+s}{2}} P_{\theta_1}[f(y_t, x_t, \theta_0) < 0]^{\frac{1-s}{2}}. \tag{A.9}$$

As f is continuously differentiable, the mean value theorem entails

$$f(y_t, x_t, \theta_1) = f(y_t, x_t, \theta_0) + \left.\frac{\partial f(y_t, x_t, \theta)}{\partial \theta'}\right|_{\theta=\overline{\theta}_t} (\theta_1 - \theta_0),\ t = 1, \ldots, n, \tag{A.10}$$

where $\overline{\theta}_t = p_t\theta_0 + (1 - p_t)\theta_1$ with $p_t = p_t(y_t, x_t, \theta_0, \theta_1) \in [0, 1],\ t = 1, \ldots, n$. Let us denote

$$H_t(\overline{\theta}_t) = \left.\frac{\partial f(y_t, x_t, \theta)}{\partial \theta'}\right|_{\theta=\overline{\theta}_t},\ t = 1, \ldots, n. \tag{A.11}$$

We can rewrite

$$f(y_t, x_t, \theta_0) = f(y_t, x_t, \theta_1) - [H_t(\overline{\theta}_t) - EH_t(\overline{\theta}_t)](\theta_1 - \theta_0) - EH_t(\overline{\theta}_t)(\theta_1 - \theta_0). \tag{A.12}$$

This yields, using equation (A.9)

$$\begin{aligned} P_{\theta_1}[s_t = s] \;&=\; P_{\theta_1}\left[u_t - \big(H_t(\overline{\theta}_t) - EH_t(\overline{\theta}_t)\big)(\theta_1 - \theta_0) > EH_t(\overline{\theta}_t)(\theta_1 - \theta_0)\right]^{\frac{1+s}{2}} \\ &\times\; P_{\theta_1}\left[u_t - \big(H_t(\overline{\theta}_t) - EH_t(\overline{\theta}_t)\big)(\theta_1 - \theta_0) \leq EH_t(\overline{\theta}_t)(\theta_1 - \theta_0)\right]^{\frac{1-s}{2}}. \end{aligned}$$

As the observations are $i.i.d.$, we will not write the subscript t. Let us denote

$$G_n^{\overline{\theta}_n}(x) = P_{\theta_n}\left[u - \big(H(\overline{\theta}_n) - EH(\overline{\theta}_n)\big)(\theta_n - \theta_0) \leq x\right] \tag{A.13}$$

where the *real* random variable $u \sim G$. Equation (A.13) can alternatively be written

$$P_{\theta_n}[s = s_a] = \left[\frac{1}{2} - G_n^{\overline{\theta}_n}\big(EH(\overline{\theta}_n)(\theta_n - \theta_0)\big)\right] s_a + \frac{1}{2} \tag{A.14}$$

where again s stands for a *real* random variable and not for a vector.
Let us now examine

$$\begin{aligned} R &= G_n^{\overline{\theta}_n}\big(EH(\overline{\theta}_n)'(\theta_n - \theta_0)\big) - G\big(EH(\overline{\theta}_n)'(\theta_n - \theta_0)\big) & \text{(A.15)} \\ &+ G\big(EH(\overline{\theta}_n)'(\theta_n - \theta_0)\big) - G(0) - G'(0)EH(\overline{\theta}_n)'(\theta_n - \theta_0) \\ &- \frac{1}{2}G''(0)(\theta_n - \theta_0)'EH(\overline{\theta}_n)EH(\overline{\theta}_n)'(\theta_n - \theta_0). & \text{(A.16)} \end{aligned}$$

When $\theta_n \to \theta_0$, we want to show that R is $o(||\theta_0 - \theta_n||^2)$. For this, we denote:

$$A = G_n^{\overline{\theta}_n}\big(EH(\overline{\theta}_n)'(\theta_n - \theta_0)\big) - G\big(EH(\overline{\theta}_n)'(\theta_n - \theta_0)\big),$$

$$\begin{aligned} B &= G\big(EH(\overline{\theta}_n)'(\theta_n - \theta_0)\big) - G(0) - G'(0)EH(\overline{\theta}_n)'(\theta_n - \theta_0) \\ &- \frac{1}{2}G''(0)(\theta_n - \theta_0)'EH(\overline{\theta}_n)EH(\overline{\theta}_n)'(\theta_n - \theta_0). \end{aligned}$$

We first consider B. We easily have

$$||B|| = o(||\theta_n - \theta_0||^2) \tag{A.17}$$

using a Taylor expansion of G in the vicinity of zero, because $EH(\overline{\theta}_n)$ is uniformly bounded by M_1 around θ_0 (condition **c**). Let us consider now A. We can major $||A||$ by

$$||A|| \leq M_1||\theta_n - \theta_0|| \sup_{y \in B(0, M_1||\theta_0 - \theta_n||)} ||G(y) - G_n(y)||. \tag{A.18}$$

Moreover, as $\{G_n\}_{n\in\mathbb{N}}$ are increasing continuous functions that converge everywhere to G, a Dini-type theorem implies the convergence is uniform. Hence, Lemma **A.1** applies. Finally

$$\sup_{y \in B(0, M_1||\theta_0 - \theta_n||)} ||G(y) - G_n(y)|| = o(||\theta_n - \theta_0||). \tag{A.19}$$

Finally

$$||A|| = o(||\theta_n - \theta_0||^2). \tag{A.20}$$

Consequently, inequalities (A.20) and (A.17) with condition **d** entail:

$$P_{\theta_n}[s_t = s] - \frac{1}{2} = s\big[-G'(0)\big(EH_t(\theta_n)'(\theta_n - \theta_0)\big) + o(||\theta_n - \theta_0||^2)\big]. \tag{A.21}$$

As $(s_1, \ldots, s_n)$ are $i.i.d.$, it follows

$$P_{\theta_n}[S = (s_1, \ldots, s_n)] = \left(\frac{1}{2}\right)^n - \left(\frac{1}{2}\right)^{n-1} \sum_{t=1}^{n} s_t\big[G'(0)\big(EH_t(\theta_n)'(\theta_n - \theta_0)\big)\big]$$

$$- \left(\frac{1}{2}\right)^{n-2} \sum_{t \leq l} s_t s_l [G'(0)^2 (\theta_n - \theta_0)(EH_t(\theta_n)')(EH_l(\theta_n))(\theta_n - \theta_0')]$$
$$+ \quad o(||\theta_n - \theta_0||^2)]. \tag{A.22}$$

The remainder follows the proof of Proposition 4.1 in Coudin and Dufour (2009) and Boldin, Simonova, and Tyurin (1997). We consider sign-based tests that maximize the mean curvature around θ_0. It is trivial to see that the locally optimal test with critical region W_α is locally unbiased (assuming the opposite goes to a contradiction), $i.e.$

$$\left.\frac{d\mathsf{P}_\theta[W_\alpha]}{d\theta}\right|_{\theta=\theta_0} = 0. \tag{A.23}$$

The behavior of the power function around zero is then totally defined by the quadratic term of its Taylor expansion which can be identified thanks to equation (A.22). The mean curvature is by definition proportional to the trace of $\frac{d^2\mathsf{P}_\theta[W]}{d\theta^2}$ at $\theta = \theta_0$ [see Boldin, Simonova, and Tyurin (1997), p. 41, Dubrovin, Fomenko, and Novikov (1984), Chapter 2, pp. 76-86 or Gray (1998), Chapter 21, pp. 373-380]. Taking the trace in the expression of equation (A.22), we find (after some computations) it is proportional to

$$\sum_{1 \leq t \neq} \sum_{l \leq n} G'(0)^2 s_t s_l EH_t(\theta_0) EH_l(\theta_0)'. \tag{A.24}$$

By adding the quantity $\sum_{t=1}^{n} \left(EH_t(\theta_n)EH_t(\theta_n)'\right)$ to (A.24), we find the locally optimal sign-based test in the sense proposed by Boldin, Simonova, and Tyurin (1997) is

$$W = \left\{ s : s'(y)\big[EH(\theta_0)EH(\theta_0)\big]' s(y) > c'_\alpha \right\} . \tag{A.25}$$

Standardizing by $EH(\theta_0)'EH(\theta_0)$ then leads to

$$W = \{ s : s'(y)EH(\theta_0)[EH(\theta_0)'EH(\theta_0)]^{-1}EH(\theta_0)'s(y) > c'_\alpha \} . \tag{A.26}$$

A.5. Proof of Theorem 7.1 (Consistency)

Consistency of IV sign-estimators is an extension of consistency of classical sign estimators [Theorem 5.9 in Coudin and Dufour (2008)]. Both proofs follow the same classical 4 steps (pointwise convergence, weak uniform convergence, consistency and identification). Here, we indicate only points that differ. The stochastic process considered here is $W^v = \{W_t^v = (y_t,\ x_t',\ v_t')\}_{t=1,2,\dots} : \Omega \to \mathbb{R}^{p+k+l}$, and we denote

$$q_t(w_t, \theta) = s\big(f(y_t, x_t, \theta)\big) \otimes h_t(v_t, \theta),\ t = 1, \dots, n, \tag{A.27}$$

which satisfies the same mixing condition. Similarly to Theorem 5.9 in Coudin and Dufour (2008), pointwise convergence for any θ is implied by assumptions A7, A9 (boundedness point) and Corollary 3.48 of White (2001).

Uniform convergence and continuity of the limiting function are implied by the generic law of large number of Andrews (1987). Andrew's conditions B1, B2 and A1 are fulfilled by assumptions A7, A8, A9 and A10. Furthermore, we use his comment 3 to conclude on the weak continuity condition (A6). Condition A6(a) allows $q_t(w,\theta)$ to have isolated discontinuities provided $q_t(w,\theta)p_t(w)$ is continuous in θ uniformly in t a.e.$[\mu]$, where μ is a σ-finite measure, that dominates each of the marginal distribution of $W_t,\ t = 1, 2\ldots$ and $p_t(w)$ is the density of W_t w.r.t. μ. Condition A6(b) states that $\int \sup_{t\geq 1}|q_t(w,\theta)|p_t(w)d\mu(w) < \infty$.
Here, we consider $\mu = P$, $q_t(w,\theta)p_t(w)$ is continuous in θ a.e. w.r.t. P, as p_t does not depend on θ and q_t is a continuous function everywhere except at $\{f(y_t,x_t,\theta) = 0\}$ which is a P-negligible set: $P[\{w\ :\ f(y_t,x_t,\theta) = 0\}] = 0$ (no tie assumption A8). Furthermore, q_t is L_1-bounded and uniformly integrable. Then, condition A6 is fulfilled. The consistency part applies without further modifications. Finally, the identification conditions A11 and A12 allow to conclude on consistency.

A.6. Proof of Theorem 7.2 (Asymptotic Normality)

If $z_t = h_t(\theta, v_t)$, Assumptions A9, A17, A14 and A15 allow to differentiate below the integral.

$$\begin{aligned}\frac{\partial}{\partial\theta'}E[h_t(\theta,v_t)s(f(y_t,x_t,\theta))] \quad &= \quad E\left[h_t(\theta,v_t)\frac{\partial f(y_t,x_t,\theta)}{\partial\theta'}g_t\big(f(y_t,x_t,\theta)|z_1,\ldots,z_n\big)\right] \\ &\quad + E\left[\frac{\partial h_t(\theta,v_t)}{\partial\theta'}\right]s\left(f(y_t,x_t,\theta)\right). \qquad \text{(A.28)}\end{aligned}$$

By uniform convergence (shown in the consistency part), it follows that the limiting objective function, $\lim_n \frac{1}{n}\sum_{t=0}^{n} E\big[z_t(\theta)s\big(f(y_t,x_t,\theta)\big)\big]$, is differentiable with derivative $L(\theta)$:

$$\begin{aligned}L(\theta) \quad &= \quad \lim_{n\to\infty}\frac{1}{n}\sum_t E\left[h_t(\theta,v_t)\frac{\partial f(y_t,x_t,\theta)}{\partial\theta'}g_t\big(f(y_t,x_t,\theta)|z_1,\ldots,z_n\big)\right] \\ &\quad + \frac{1}{n}\sum_t E\left[\frac{\partial h_t}{\partial\theta'}s\big(f(y_t,x_t,\theta)\big)\right].\end{aligned}$$

Theorem 7.2 in Newey and McFadden (1994) may then be applied. Their condition (i), which states that 0 is attained at the limit by θ_0, is fulfilled by the moment condition A3. Their condition (ii) states that the limit objective function is differentiable at θ_0 and positive definite. This is fulfilled by the first part of our proof and condition A19. Then, their condition (iii) (interior) is implied by A10. Using the mixing specification A16 of $\{w\}$ and conditions A3, A9, A13 and A18, we apply a White-Domowitz central limit theorem [see White (2001), Theorem 5.20]. This fulfills condition (iv) of Theorem 7.2 in Newey and McFadden (1994). Finally, condition v (stochastic equicontinuity) is implied by uniform convergence (see the consistency part) which completes the proof.

References

AMEMIYA, T. (1982): "Two-Stage Least Absolute Deviations Estimator," *Econometrica*, 50(3), 689–711.

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(1), 46–63.

ANDREWS, D., AND V. MARMER (2008): "Exactly Distribution-Free Inference in Instrumental Variables Regression with Possible Weak Instruments," *Journal of Econometrics*, 142(1), 183–200.

ANDREWS, D., M. MOREIRA, AND J. STOCK (2004): "Optimal Invariant Similar Tests for Instrumental Variables Regression," Discussion paper T0299, N.B.E.R.

——— (2007): "Performance of Conditional Wald Tests in IV Regression with Weak Instruments," *Journal of Econometrics*, 139(1), 116–132.

ANDREWS, D., AND J. STOCK (2005): "Inference with Weak Instruments," Cowles Foundation Discussion papers 1530, Cowles Foundation for Research Economics, Yale University.

ANDREWS, D. W. (1987): "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," *Econometrica*, 55(6), 1465–1471.

ANGRIST, J., AND A. KRUEGER (1991): "Does Compulsory School Attendance Affect Schooling and Earning?," *The Quarterly Journal of Economics*, 106(4), 979–1014.

——— (1995): "Split-Sample Instrumental Variables Estimates of the Return to Schooling," *Journal of Business and Economic Statistics*, 13(2), 225–235.

BARNARD, G. A. (1963): "Comment on 'The Spectral Analysis of Point Processes' by M. S. Bartlett," *Journal of the Royal Statistical Society, Series B*, 25(2), 294.

BEKKER, P., AND F. KLEIBERGEN (2003): "Finite-Sample Instrumental Variables Inference Using an Asymptoticall Pivotal Statistic," *Econometric Theory*, 19(1), 744–753.

BEKKER, P., AND S. LAWFORD (2008): "Symmetry-Based Inference in an Instrumental Setting," *Journal of Econometrics*, 142(1), 28–49.

BOLDIN, M. V., G. I. SIMONOVA, AND Y. N. TYURIN (1997): *Sign-Based Methods in Linear Statistical Models*, vol. 162 of *Translations of Mathematical Monographs*. American Mathematical Society, Maryland.

BOUND, J., D. JAEGER, AND R. BAKER (1995): "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variables is Weak," *Journal of the American Statistical Association*, 90(430), 443–450.

BUCHINSKY, M. (1998): "Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research," *Journal of Human Resources*, 33(1), 88–126.

CAMPBELL, B., AND J.-M. DUFOUR (1991): "Over-rejections in Rational Expectations Models: A Nonparametric Approach to the Mankiw-Shapiro Problem," *Economics Letters*, 35(3), 285–290.

——— (1995): "Exact Nonparametric Orthogonality and Random Walk Tests," *Review of Economics and Statistics*, 77(1), 1–16.

CHEN, L., AND S. PORTNOY (1996): "Two-Stage Regression Quantiles and Two-Stage Trimmed Least Squares Estimators for Structural Equation Models," *Communications in Statistics- Theory and Methods*, 25, 1005–1032.

CHERNOZHUKOV, V., AND C. HANSEN (2008): "Instrumental variable quantile regression: A robust inference approach," *Journal of Econometrics*, 142(1), 379–398.

CHERNOZHUKOV, V., C. HANSEN, AND M. JANSSON (2009): "Finite Sample Inference for Quantile Regression Models," *Journal of Econometrics*, 152(2), 93–103.

CHESHER, A. (2005): "Nonparametric Identification under Discrete Variation," *Econometrica*, 73(5), 1525–1550.

COUDIN, E., AND J.-M. DUFOUR (2008): "Hodges-Lehmann Sign-based Estimators and Generalized Confidence Distributions in Linear Median Regressions with Heterogenous Dependent Errors," Discussion Paper 2008-33, CREST.

COUDIN, E., AND J.-M. DUFOUR (2009): "Finite-Sample Distribution-Free Inference in Linear Median Regressions under Heteroskedasticity and Nonlinear Dependence of Unknown Form," *Econometrics Journal, 10th anniversary special edition*, 12(S1), S19–S49.

CRUZ, L., AND M. MOREIRA (2005): "On the Validity of Econometric Techniques with Weak Instruments," *The Journal Of Human Resources*, 40(2), 393–410.

DAVIDSON, R., AND J. MCKINNON (2001): "Artificial regressions," Discussion paper 1038, Queen's University, Department of Economics.

DUBROVIN, B., A. FOMENKO, AND S. NOVIKOV (1984): *Modern-Geometry- Methods and Applications*, Graduate texts in Mathematics. Springer-Verlag, New-York.

DUFOUR, J.-M. (1981): "Rank Tests for Serial Dependence," *Journal of Time Series Analysis*, 2(3), 117–128.

——— (1990): "Exact Tests and Confidence Sets in Linear Regressions with Autocorrelated Errors," *Econometrica*, 58(2), 475–494.

——— (1997): "Some Impossibility Theorems in Econometrics, with Applications to Structural and Dynamic Models," *Econometrica*, 65(6), 1365–1389.

——— (2003): "Identification, Weak Instruments and Statistical Inference in Econometrics. Presidential Address to the Canadian Economics Association," *Canadian Journal of Economics*, 36(4), 767–808.

——— (2006): "Monte Carlo Tests with Nuisance Parameters: A General Approach to Finite-Sample Inference and Nonstandard Asymptotics in Econometrics," *Journal of Econometrics*, 133(2), 443–477.

DUFOUR, J.-M., AND J. JASIAK (2001): "Finite Sample Limited Information Inference Methods for Structural Equations and Models with Generated Regressors," *International Economic Review*, 42(3), 815–843.

DUFOUR, J.-M., AND J. KIVIET (1998): "Exact Inference Methods for First-order Autoregressive Distributed Lag Models," *Econometrica*, 82(1), 79–104.

DUFOUR, J.-M., AND A. TAAMOUTI (2010): "Exact Optimal and Adaptative Inference in Linear and Nonlinear Models Under Heteroskedasticity and Non-Normality of Unknown Forms," *Computational Statistics and Data Analysis*, 54(11), 2532–2553, The Fifth Special Issue on Computational Econometrics.

DUFOUR, J.-M., AND M. TAAMOUTI (2002): "Point-Optimal Instruments and Generalized Anderson-Rubin procedures for Nonlinear models," Discussion paper, C.R.D.E., Université de Montréal.

——— (2005): "Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments," *Econometrica*, 73(4), 1351–1365.

DWASS, M. (1957): "Modified Randomization Tests for Nonparametric Hypotheses," *Annals of Mathematical Statistics*, 28(1), 181–187.

GOFFE, W. L., G. D. FERRIER, AND J. ROGERS (1994): "Global Optimization of Statistical Functions with Simulated Annealing," *Journal of Econometrics*, 60(1), 65–100.

GOURIÉROUX, C., AND A. MONFORT (1995): *Statistics and Econometric Models*, vol. 2. Cambridge University Press.

GRAY, A. (1998): *Modern Differential Geometry of Curves and Surfaces with Mathematica*. CRC Press, Boca Raton, Florida, 2nd edn.

HANSEN, C., J. HAUSMAN, AND W. NEWEY (2005): "Estimation with Many Instrumental Variables," Discussion paper, M.I.T., Department of Economics.

HAUSMAN, J. (1983): "Specification and Estimation of Simultaneous Models," in *Handbook of Econometrics*, ed. by Z. Grilichest, and M. Intriligator, vol. 1, chap. 7, pp. 391–448. Marcel Dekker, New York.

HODGES, J. L., AND E. L. LEHMANN (1963): "Estimates of Location Based on Rank Tests," *The Annals of Mathematical Statistics*, 34(2), 598–611.

HONG, H., AND E. TAMER (2003): "Inference in Censored Models With Endogenous Regressors," *Econometrica*, 71(3), 905–932.

HONORE, B., AND L. HU (2004): "On the Performance of Some Robust Instrumental Variables," *Journal of Business & Economic Statistics*, 22(1), 30–39.

KLEIBERGEN, F. (2002): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70(5), 1781–1803.

——— (2005): "Testing Parameters in GMM Without Assuming They Are Identified," *Econometrica*, 73(4), 1103–1123.

——— (2007): "Generalizing Weak Instrument Robust IV Statistics Towards Multiple Parameters, Unrestricted Covariance Matrices and Identification Statistics," *Journal of Econometrics*, 139(1), 181–216.

LEE, S. (2003): "Efficient Semiparametric Estimation of a Partially Linear Quantile Regression Model," *Econometric Theory*, 19(1), 1–31.

LEHMANN, E. L., AND C. STEIN (1949): "On the Theory of some Non-Parametric Hypotheses," *Annals of Mathematical Statistics*, 20(1), 28–45.

MOREIRA, M. (2001): "Tests with Correct Size When Instruments can be Arbitrary Weak," Discussion paper 37, Center for Labor Economics, UC Berkeley.

——— (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71(4), 1027–1048.

MOREIRA, M., AND B. POI (2003): "Implementing Conditional Tests with Correct Size in the Simultaneous Equations Model," *The Stata Journal*, 3(1), 57–70.

NELSON, C., R. STARTZ, AND E. ZIVOT (1998): "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," *International Economic Review*, 39(4), 1119–1144, Symposium on Forcasting and Empirical Methods in Macroeconomics and Finance.

NEWEY, W. K., AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, chap. 36, pp. 2113–2245. Elsevier, North Holland.

POWELL, J. (1983): "The Asymptotic Normality of Two-stage Least Absolute Deviations Estimators," *Econometrica*, 51(5), 1569–1576.

PRESS, W., S. TEUKOLSKY, W. VETTERLING, AND B. FLANNERY (2002): *Numerical Recipes in Fortran 90*. Press Syndicate of the University of Cambridge, USA.

SAKATA, S. (2001): "Instrumental Variable Estimation Based on Mean Absolute Deviation," Discussion paper, Michigan - Center for research on economics and social theory, First version 1998, revised 2001.

SO, B. S. (1994): "A Sharp Cramér-Rao Type Lower-Bound for Median-Unbiased Estimators," *Journal of the Korean Statistical Society*, 23(1).

STAIGER, D., AND J. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65(3), 557–586.

STOCK, J., AND J. WRIGHT (2000): "GMM with Weak Identification," *Econometrica*, 68(5), 1055–1096.

SUNG, N. K., G. STANGENHAUS, AND H. DAVID (1990): "A Cramér-Rao Analogue for Median-Unbiased Estimators," *Trabajos de Estadistica*, 5, 83–94.

TIPPETT, L. (1931): *The Methods of Statistics*. Williams and Norgate, London.

WANG, J., AND E. ZIVOT (1998): "Inference on Structural Parameters in Instrumental Variables Regression with weak instruments," *Econometrica*, 66(6), 1389–1404.

WHITE, H. (2001): *Asymptotic Theory For Econometricians*. Academic Press, New York, revised version.

WRIGHT, J. H. (2000): "Alternative Variance-Ratio Tests Using Ranks and Signs," *Journal of Business and Economic Statistics*, 18(1), 1–9.