Finite-sample generalized confidence distributions and sign-based robust estimators in median regressions with heterogeneous dependent errors [*]

Élise Coudin [†]
INSEE-CREST

Jean-Marie Dufour [‡]
McGill University

January 2017

[*] The authors thank Magali Beffy, Marine Carrasco, Frédéric Jouneau, Marc Hallin, Thierry Magnac, Bill McCausland, Benoit Perron, and Alain Trognon for useful comments and constructive discussions. This work was supported by the William Dow Chair in Political Economy (McGill University), the Bank of Canada (Research Fellowship), the Toulouse School of Economics (Pierre-de-Fermat Chair of excellence), the Universitad Carlos III de Madrid (Banco Santander de Madrid Chair of excellence), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds de recherche sur la société et la culture (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. TEL: 33 (0) 1 41 17 77 33; FAX: 33 (0) 1 41 17 76 34.

[‡] William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 414, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 4400 ext. 09156; FAX: (1) 514 398 4800; e-mail: `jean-marie.dufour@mcgill.ca` . Web page: `http://www.jeanmariedufour.com`

ABSTRACT

We study the problem of estimating the parameters of a linear median regression without any assumption on the shape of the error distribution – including no condition on the existence of moments – allowing for heterogeneity (or heteroskedasticity) of unknown form, noncontinuous distributions, and very general serial dependence (linear and nonlinear). This is done through a *reverse inference approach*, based on a distribution-free testing theory [Coudin and Dufour (2009, *The Econometrics Journal*)], from which confidence sets and point estimators are subsequently generated. The estimation problem is tackled in two complementary ways. *First*, we show how *confidence distributions* for model parameters can be applied in such a context. Such distributions – which can be interpreted as a form of *fiducial inference* – provide a frequency-based method for associating probabilities with subsets of the parameter space (like posterior distributions do in a Bayesian setup) without the introduction of prior distributions. We consider *generalized confidence distributions* applicable to multidimensional parameters, and we suggest the use of a projection technique for confidence inference on individual model parameters. *Second*, we propose point estimators, which have a natural association with confidence distributions. These estimators are based on maximizing test p-values and inherit robustness properties from the generating distribution-free tests. Both finite-sample and large-sample properties of the proposed estimators are established under weak regularity conditions. We show they are median unbiased (under symmetry and estimator unicity) and possess equivariance properties. Consistency and asymptotic normality are established without any moment existence assumption on the errors, allowing for noncontinuous distributions, heterogeneity and serial dependence of unknown form. These conditions are considerably weaker than those used to show corresponding results for LAD estimators. In a Monte Carlo study of bias and RMSE, we show sign-based estimators perform better than LAD-type estimators in heteroskedastic settings. We present two empirical applications, which involve financial and macroeconomic data, both affected by heavy tails (non-normality) and heteroskedasticity: a trend model for the S&P index, and an equation used to study β-convergence of output levels across U.S. States.

Key words: sign-based methods; median regression; test inversion; Hodges-Lehmann estimators; confidence distributions; p-value function; least absolute deviation estimators; quantile regressions; sign test; simultaneous inference; Monte Carlo tests; projection methods; non-normality; heteroskedasticity; serial dependence; GARCH; stochastic volatility.
Journal of Economic Literature classification: C13, C12, C14, C15.

Contents

List of Definitions, Assumptions, Propositions and Theorems

List of Tables

1. Introduction

A basic problem in statistics and econometrics consists in studying the relationship between a dependent variable and a vector of explanatory variables under weak distributional assumptions. For that purpose, the Laplace-Boscovich median regression is an attractive approach because it can yield estimators and tests which are considerably more robust to non-normality and outliers than least-squares methods; see Dodge (1997). The least absolute deviation (LAD) estimator is the reference estimation method in this context. Quantile regressions [Koenker and Bassett (1978), Koenker (2005)] can be viewed as extensions of median regression. An important reason why such methods yield more robust inference comes from the fact that hypotheses about moments are not generally testable in nonparametric setups, while hypotheses about quantiles remain testable under similar conditions [see Bahadur and Savage (1956), Dufour (2003), Dufour, Jouneau and Torrès (2008)].

Tests and confidence sets associated with LAD estimators are typically based on asymptotic approximations. The distributional theory of LAD estimators and their extensions usually postulates moment conditions on model errors, such as the existence of moments up to a given order, as well as other regularity conditions, such as continuity, independence or identical distributions; see for instance Portnoy (1991), Knight (1998), El Bantli and Hallin (1999), and Koenker (2005). Further, this theory and the associated tests and confidence sets are typically based on asymptotic approximations. The same remark applies to work on LAD-type estimation in models involving heteroskedasticity and autocorrelation [Zhao (2001), Weiss (1990)], endogeneity [Amemiya (1982), Powell (1983), Hong and Tamer (2003)], censored models [Powell (1984, 1986)], and nonlinear functional forms [Weiss (1991)]. By contrast, provably valid tests can be derived in such models, under remarkably weaker conditions, which do not require the existence of moments and allow for very general forms of heterogeneity (or heteroskedasticity); see Coudin and Dufour (2009).

In this paper, we exploit this feature of testing theory in the context of median regression to derive more robust estimation methods. Specifically, we study the problem of estimating the parameters of a linear median regression without any assumption on the shape of the error distribution – including no condition on the existence of moments at any order – allowing for heterogeneity (or heteroskedasticity) of unknown form, noncontinuous distributions, and very general serial dependence (linear and nonlinear). This is done through a *reverse inference approach*, which starts from a distribution-free testing theory [Coudin and Dufour (2009)], subsequently exploited to derive confidence sets and point estimators. Using the tests proposed in Coudin and Dufour (2009), the estimation problem is tackled in two complementary ways.

First, we show how *confidence distributions* for model parameters [Schweder and Hjort (2002), Xie and Singh (2013)] can be applied in such a context. Such distributions – which can be interpreted as a form of *fiducial inference* [Fisher (1930), Buehler (1983), Efron (1998)] – provide a frequency-based method for associating probabilities with subsets of the parameter space (like posterior distributions do in a Bayesian setup) without the introduction of a prior distribution. In the one-dimensional model, the confidence distribution is defined as a distribution whose quantiles span all the possible confidence intervals [Schweder and Hjort (2002)]. In this paper, we consider *generalized confidence distributions* applicable to multidimensional parameters, and we suggest the use of a projection technique for confidence inference on individual model parameters. The latter are exact – in the sense that the parameters considered are covered with known probabilities (or larger) – under the mediangale assumption considered in Coudin and Dufour (2009). Further, if more general linear dependence is allowed, the proposed method remains valid asymptotically.

Second, we propose point estimates, which bear a natural association with the above confidence distributions. These Hodges-Lehmann estimators are based on maximizing test p-values and inherit several robustness properties from the distribution-free tests used to generate them [Hodges and Lehmann (1963)]. In particular, both finite-sample and large-sample properties are established under very weak regularity conditions. We show they are median unbiased (under symmetry and estimator unicity) and possess equivariance properties with respect to linear transformations of model variables. Consistency and asymptotic normality are established without any moment existence assumption on the errors, allowing noncontinuous distributions, heterogeneity and general serial dependence of unknown form. These conditions are considerably weaker than those usually used to obtain corresponding results for LAD estimators.

The conjunction of sign-based tests, projection-based confidence regions, projection-based p-values and sign-based estimators thus provides a complete system of inference, which is valid for any given sample size under very weak distributional assumptions and remains asymptotically valid under weaker conditions (including allowance for general forms of linear residual dependence).

Fisher's fiducial distributions and other fiducial inference arguments [Fisher (1930), Buehler (1983), Efron (1998), Hannig (2006)] are not commonly used in econometrics because they require the availability of pivotal test statistics with known distributions. This condition is not fulfilled in general, especially in semiparametric or nonparametric settings. However, in the context of median regression, sign-based methods provide a way to construct such pivots and fiducial inference tools can be developed. For any given sample size, the sign transform enables one to construct test statistics with known nuisance-parameter-free distributions without additional parametric restrictions. This enables us to construct fiducial inference tools adapted to multidimensional parameters. We exploit realized p-value functions, which are constructed by testing hypotheses of the form $H_0(\beta_0) : \beta = \beta_0$, where β is the vector of the regression coefficients. Specifically, we combine sign-based tests for such joint hypotheses [as given in Coudin and Dufour (2009)] with projection techniques. For each component, a projected p-value function provides a representation of the evidence for each possible value of that component.

Using the above p-values (as a function of β_0), we then derive estimators and study their properties. Hodges and Lehmann (1963) proposed a general principle to directly derive estimators from test procedures. They suggest to invert a test for $H_0(\beta_0) : \beta = \beta_0$, and then choose the value β_0 which is "least rejected" by the test procedure. First applied to the Wilcoxon's signed rank-statistic for estimating a shift or a location, this principle was adapted to regression models by deriving so-called R-estimators from rank or signed-rank statistics [Jureckova (1971), Jaeckel (1972), Koul (1971)]. In a multidimensional context, this leads one to select the value of β_0 with the highest *degree of confidence, i.e.* with the highest p-value.

We study the problem of estimating the parameters of the median regression by minimizing (weighted) sign-based test statistics over different null hypotheses. Since the tests used to generate them are remarkably robust, the estimators inherit strong robustness properties. Sign-based estimators are generated by sign-based tests, and they inherit several attractive properties of the latter (*e.g.*, robustness to non-normality and heterogeneity). We will see that these can alternatively be computed by minimizing quadratic forms of the constrained signs. So they have a classical GMM form [Hansen (1982), and Honore and Hu (2004) for GMM statistics involving signs].

Both finite-sample and large-sample properties of sign-based estimators are established under weak regularity conditions. We show they are median unbiased (under symmetry and estimator unicity) and possess equivariance properties with respect to linear transformations of model variables. Consistency and asymptotic normality are established without any moment existence assumption on the errors, allowing noncontinuous distributions, heterogeneity and general serial dependence of unknown form. These conditions are considerably weaker than those usually used to obtain corresponding results for LAD estimators; see Bassett and Koenker (1978), Bloomfield and Steiger (1983), Powell (1984), Phillips (1991), Pollard (1991), Portnoy (1991), Weiss (1991), Knight (1998), El Bantli and Hallin (1999) and the references therein. In particular, asymptotic normality and consistency hold for heavy-tailed disturbances which may not have finite variances. This interesting property is induced by the sign transformation. Signs of residuals always possess finite moments, so no further restriction on the disturbance moments is required. Except for Knight (1989) and Phillips (1991), who considered the case of autoregressive models, the distribution of LAD estimators in regressions where the error variances may not exist has received little attention. In general, LAD estimators and the sign-based estimators proposed here follow from different optimization rules, and they can be quite different.

The class of sign-based estimators we propose includes as special cases the *sign estimators* derived by Boldin, Simonova and Tyurin (1997) from locally most powerful sign tests in linear regressions with *i.i.d.* errors and fixed regressors. Note also that the procedures proposed by Hong and Tamer (2003) and Honore and Hu (2004) also rely on the *i.i.d.* assumption. In this paper, we stress that a major advantage of signs over ranks consists in dealing transparently with heteroskedastic (or heterogeneous) disturbances. Many heteroskedastic and possibly dependent schemes are covered and, in presence of linear dependence, a HAC-type correction for heteroskedasticity and autocorrelation can be included in the criterion function.

The construction of sign-based estimators as Hodges-Lehmann estimators makes these a natural complement of the finite-sample tests used to generate them. The latter rely on the exact distribution of the corresponding sign-based test statistics, do not involve nuisance parameters, and allow one to control test levels in finite samples under heteroskedasticity and nonlinear dependence of unknown form. In Coudin and Dufour (2009), Monte Carlo test methods [Dwass (1957), Barnard (1963) and Dufour (2006)] are combined with test inversion and projection techniques [Dufour (1990, 1997), Dufour and Kiviet (1998), Abdelkhalek and Dufour (1998), Dufour and Jasiak (2001), Dufour and Taamouti (2005)] to build confidence sets and test general hypotheses.[1] There is no need to estimate the error density at zero in contrast with tests that rely on kernel estimates of the LAD asymptotic covariance matrix.[2] Furthermore, when the test criteria are modified to cover linear dependence,

[1]For an alternative finite-sample inference exploiting a quantile version of the same sign pivotality result, which holds if the observations are X-conditionally independent, see Chernozhukov, Hansen and Jansson (2009).

[2]In the *i.i.d.* error case, Honore and Hu (2004) observed in simulations that kernel-based estimates of the asymptotic standard error of the median-based estimator tend to be too small, so the associated tests tend to overreject the null hypothesis. Other estimates of the LAD asymptotic covariance matrix can be obtained by bootstrap procedures [design matrix bootstrap in Buchinsky (1995, 1998), block bootstrap in Fitzenberger (1997), Bayesian bootstrap in Hahn (1997)]

the resulting inference is asymptotically valid. The conjunction of sign-based tests, projection-based confidence regions, and sign-based estimators thus provides a complete system of inference, which is valid for any given sample size under very weak distributional assumptions and remains asymptotically valid under even weaker conditions (including allowance for linear dependence in regression disturbances).

We study the performance of the proposed estimators in a Monte Carlo study that allows for various non-Gaussian and heteroskedastic setups. We find that sign-based estimators are competitive (in terms of bias and RMSE) when errors are *i.i.d.*, while they are substantially more reliable than usual methods (LS, LAD) when heterogeneity or serial dependence is present in the error term.

Finally we present two empirical applications, which involve financial and macroeconomic data. In the first one, we study a trend model for the Standard and Poor's Composite Price Index, over the period 1928-1987 as well as the 1929 crash period (which is characterized by huge price volatilities). In the second application, we consider an equation used to study β-convergence of output levels across U.S. States, with a small size. In both cases, the data are affected by heavy tails (non-normality) and heteroskedasticity.

The paper is organized as follows. Section 2 presents the model, the sign-based statistics and the Monte Carlo tests. Section 3 is dedicated to confidence distributions and p-value functions. In section 4, we define the proposed family of sign-based estimators. The finite-sample properties of the sign-based estimators are studied in section 5, while their asymptotic properties are considered in section 6. In section 7, we present the results of our simulation study of bias and RMSE. The empirical application is reported in section 8. We conclude in section 9. Appendix A contains the proofs.

2. Framework

We will now summarize the general framework we study and define the test statistics on which the estimation methods we propose are based. This framework is the same as the one used in Coudin and Dufour (2009).

2.1. Model

We consider a stochastic process $\{(y_t, x_t') : \Omega \to \mathbb{R}^{p+1} : t = 1, 2, \dots\}$ defined on a probability space $(\Omega, \mathscr{F}, \mathsf{P}),$ such that y_t and x_t satisfy a linear model of the form

$$y_t = x_t'\beta + u_t, \quad t = 1, \dots, n, \tag{2.1}$$

where y_t is a dependent variable, $x_t = (x_{t1}, \dots, x_{tp})'$ is a p-vector of explanatory variables, and u_t is an error process. The x_t's may be random or fixed. In the sequel, $y = (y_1, \dots, y_n)' \in \mathbb{R}^n$ will denote the dependent variable vector, $X = (x_1, \dots, x_n)' \in \mathbb{R}^{n \times p}$ the $n \times p$ matrix of explanatory variables, and $u = (u_1, \dots, u_n)' \in \mathbb{R}^n$ the disturbance vector. Moreover, $F_t(\cdot\,|x_1, \dots, x_n)$ represents

and resampling methods [Parzen, Wei and Ying (1994)]. But the justification of these also rely on usual asymptotic regularity conditions.

the distribution function of u_t conditional on X. This framework is also used in Coudin and Dufour (2009).

The traditional form of a median regression assumes that the disturbances $u_1, \ldots, u_n$ are *i.i.d.* with median zero

$$\text{Med}(u_t|x_1, \ldots, x_n) = 0, \; t = 1, \ldots, n. \tag{2.2}$$

Here, we relax the assumption that the u_t are *i.i.d.*, and we consider moment conditions based on residual signs where the sign operator $s : \mathbb{R} \to \{-1, 0, 1\}$ is defined as $s(a) = \mathbf{1}_{[0,+\infty)}(a) - \mathbf{1}_{(-\infty,0]}(a)$, with $\mathbf{1}_A(a) = 1$ if $a \in A$ and $\mathbf{1}_A(a) = 0$ if $a \notin A$. For convenience, if $u \in \mathbb{R}^n$, we will note $s(u) = \big(s(u_1), \ldots, s(u_n)\big)$, the n-vector of the signs of the components.

Assumption (2.2) is not sufficient to obtain a finite-sample distributional theory for sign statistics (because further restrictions on the dependence between the errors are needed). Let us consider *adapted sequences* $\mathscr{S}(\mathbf{v}, \mathscr{F}) = \{v_t, \mathscr{F}_t : t = 1, 2, \ldots\}$ where v_t is any measurable function of $W_t = (y_t, x_t')'$, $\mathscr{F}_t$ is a σ-field in Ω, $\mathscr{F}_s \subseteq \mathscr{F}_t$ for $s < t$, $\sigma(W_1, \ldots, W_t) \subset \mathscr{F}_t$ and $\sigma(W_1, \ldots, W_t)$ is the σ-algebra spanned by $W_1, \ldots, W_t$. Then the *weak conditional mediangale* provides such a setup.

Assumption 2.1 WEAK CONDITIONAL MEDIANGALE. *Let $\mathscr{F}_t = \sigma(u_1, \ldots, u_t, X)$, for $t \geq 1$. $\mathbf{u}$ in the adapted sequence $\mathscr{S}(\mathbf{u}, \mathscr{F})$ is a weak mediangale conditional on X with respect to $\{\mathscr{F}_t : t = 1, 2, \ldots\}$ iff $\mathsf{P}[u_1 < 0|X] = \mathsf{P}[u_1 > 0|X]$ and*

$$\mathsf{P}[u_t < 0|u_1, \ldots, u_{t-1}, X] = \mathsf{P}[u_t > 0|u_1, \ldots, u_{t-1}, X], \; \textit{for } t > 1. \tag{2.3}$$

Besides nonnormality (including no condition on the existence of moments), this assumption allows for heterogeneity (or heteroskedasticity) of unknown form, noncontinuous distributions, and general forms of (nonlinear) serial dependence, including GARCH-type and stochastic volatility of unknown order. It does not, however, cover “linear serial dependence” such as an ARMA process on u_t.

Clearly, Assumption 2.1 entails (2.2). When $\mathsf{E}|x_t| < +\infty$, for all t, it also implies that $s(u_t)$ is uncorrelated with x_t, an assumption we state for future reference.

Assumption 2.2 SIGN MOMENT CONDITION. $\mathsf{E}|x_t| < +\infty$ *and* $\mathsf{E}[s(u_t)x_t] = 0$, *for* $t = 1, \ldots, n$.

This assumption allows for both linear and nonlinear serial dependence, but makes difficult the derivation of finite-sample distributions. We use it in the asymptotic results presented below.

2.2. Quadratic sign-based tests

In order to derive robust estimators, we consider tests for hypotheses of the form $H_0(\beta_0) : \beta = \beta_0$ vs. $H_1(\beta_0) : \beta \neq \beta_0$ in model (2.1)-(2.2). These are based on general quadratic forms based on the vector $s(y - X\beta_0)$ of the constrained signs (*i.e.*, the signs aligned with respect to $X\beta_0$):

$$D_S[\beta_0, \bar{\Omega}_n(\beta_0)] = s(y - X\beta_0)' X \Omega_n\big[s(y - X\beta_0), X\big] X' s(y - X\beta_0) \tag{2.4}$$

where $\bar{\Omega}_n(\beta_0) = \Omega_n\big[s(y - X\beta_0), X\big]$ is a $p \times p$ positive definite weight matrix which may depend on the constrained signs. If the disturbances follow a weak mediangale (Assumption 2.1), sign-based statistics of this form constitute pivotal functions: the distribution of $D_S[\beta_0, \bar{\Omega}_n(\beta_0)]$ conditional on X is completely determined under $H_0(\beta_0)$ and can be simulated. Even though the distribution of $D_S[\beta_0, \bar{\Omega}_n(\beta_0)]$ depends on X and $\Omega_n[\,\cdot\,]$ under $H_0(\beta_0)$, critical values can be approximated to any degree of precision by simulation. Alternatively, exact Monte Carlo tests can be built using a randomized tie correction procedure [Dufour (2006)]. So we can get an exact test of $H_0(\beta_0)$. The fact that $\Omega_n[\,\cdot\,]$ depends on the data only through $s(y-X\beta_0)$ plays a central role in generating this feature.

Further, if linear serial dependence is allowed and the assumption that $s(y-X\beta_0)$ are X are independent is relaxed [as described in Coudin and Dufour (2009)], this dependence can be taken into account by an appropriate choice of $\Omega_n[\,\cdot\,]$. The test statistic $D_S[\beta_0, \bar{\Omega}_n(\beta_0)]$ then remains asymptotically pivotal under $H_0(\beta_0)$, and the finite-sample procedure just described yields a test such that the probability of rejecting $H_0(\beta_0)$ converges to the nominal level of test under any distribution compatible with $H_0(\beta_0)$. In all cases, due to the sign transformation, the tests so obtained are remarkably robust to heavy-tailed distributions (and other features).

It will be useful to spell out how an exact Monte Carlo test based on a discrete test statistic like $D_S[\beta_0, \bar{\Omega}_n(\beta_0)]$ can be obtained. Under Assumption 2.1, we can generate a vector of N independent replicates $\big(D_S^{(1)}(\beta_0), \dots, D_S^{(N)}(\beta_0)\big)'$ from the distribution of $D_S[\beta_0, \bar{\Omega}_n(\beta_0)]$ under the null hypothesis as well as $(V^{(0)}, \dots, V^{(N)})'$ a $(N+1)$-vector of *i.i.d.* uniform variables on the interval $[0, 1]$. Setting $D_S^{(0)}(\beta_0) \equiv D_S[\beta_0, \bar{\Omega}_n(\beta_0)]$ the observed statistic. Then, a Monte Carlo test for $H_0(\beta_0)$ consists in rejecting the null hypothesis whenever the empirical p-value is smaller than α, *i.e.* $\tilde{p}_N(\beta_0) \leq \alpha$ where $\tilde{p}_N(\beta_0) \equiv \hat{p}_N[D_S^{(0)}(\beta_0), \beta_0]$,

$$\hat{p}_N(x, \beta_0) = \frac{N\hat{G}_N(x, \beta_0) + 1}{N+1} \tag{2.5}$$

and $\hat{G}_N(x, \beta_0) = 1 - \frac{1}{N}\sum_{i=1}^{N} s_+(x - D_S^{(i)}(\beta_0)) + \frac{1}{N}\sum_{i=1}^{N} \delta(D_S^{(i)}(\beta_0) - x) s_+(V^{(i)} - V^{(0)})$, with $s_+(x) = \mathbf{1}_{[0,\infty)}(x)$, $\delta(x) = \mathbf{1}_{\{0\}}(x)$. When $\alpha(N+1)$ is an integer, the size of this test is equal to α for any sample size n [see Dufour (2006)]. This procedure also provides a test such that the probability of rejection converges to α.

Note also that the confidence region

$$C_{1-\alpha}(\beta) = \{\beta_0 : \tilde{p}_N(\beta_0) \geq \alpha\} \tag{2.6}$$

which contains all the values β_0 such that the empirical p-value $\tilde{p}_N(\beta_0)$ is higher than α has by construction level $1-\alpha$ for any sample size. It is then possible to derive general (and possibly nonlinear) tests and confidence sets by projection techniques. For example, conservative individual confidence intervals are obtained in such a way. Finally, if D_S is an asymptotically pivotal function all previous results hold asymptotically. For a detailed presentation, see Coudin and Dufour (2009).

3. Confidence distributions

In the one-parameter model, statisticians have defined the confidence distribution notion that summarizes a family of confidence intervals; see Schweder and Hjort (2002). By definition, the quantiles of a confidence distribution span all the possible confidence intervals of a real β. The confidence distribution is a reinterpretation of the Fisher fiducial distributions and provides, in a sense, an analogue of Bayesian posterior probabilities in a frequentist setup [see also Fisher (1930), Neyman (1941) and Efron (1998)]. This statistical notion is not commonly used in the econometric literature, for two reasons. First, it is only defined in the one-parameter case. Second, it requires that the test statistic be a pivot with known exact distribution. Below we extend that notion (or an equivalent) to multidimensional parameters. The sign transformation enables one to construct statistics which are pivots with known distribution without imposing parametric restrictions on the sample. Consequently, our setup does not suffer from the second restriction. In that section, we briefly recall the initial statistical concept and apply it to an example in univariate regression. Then, we address the extension to multidimensional regressions.

3.1. Confidence distributions in univariate regressions

Schweder and Hjort (2002) defined the confidence distribution for the real parameter β such a distribution depending on the observations (y,x), whose cumulative distribution function evaluated at the true value of β has a uniform distribution whatever the true value of β. In a formalized way, this can be expressed as follows:

Definition 3.1 CONFIDENCE DISTRIBUTION. *Any distribution with cumulative $CD(\beta)$ and quantile function $CD^{-1}(\beta)$, such that*

$$P_\beta[\beta \le CD^{-1}(\alpha;y;x)] = P_\beta[CD(\beta;y;x) \le \alpha] = \alpha \tag{3.1}$$

for all $\alpha \in (0,1)$ and for all probability distributions in the statistical model, is called a confidence distribution of β.

$(-\infty,\ CD^{-1}(\alpha)]$ is a one-sided stochastic confidence interval with coverage probability α,[3] and the realized confidence $CD(\beta_0;y;x)$ is the p-value of the one-sided hypothesis $H_0^*(\beta_0):\ \beta \le \beta_0$ versus $H_1^*(\beta_0):\ \beta > \beta_0$ when the observed data are y, x. The realized p-value when testing $H_0(\beta_0):\ \beta = \beta_0$ versus $H_1(\beta_0):\ \beta \ne \beta_0$ is $2\min\{CD(\beta_0), 1 - CD(\beta_0)\}$. Those relations are stated in Lemma 2 of Schweder and Hjort (2002): *the confidence of the statement "$\beta \le \beta_0$" is the degree of confidence $CD(\beta_0)$ for the confidence interval $\big(-\infty, CD^{-1}\big(CD(\beta_0)\big)\big]$, and is equal to the p-value of a test of $H_0^*(\beta_0): \beta \le \beta_0$ vs. $H_1^*(\beta_0): \beta > \beta_0$*. Hence, tests and confidence intervals on β are contained in the confidence distribution.

Schweder and Hjort (2002) also note that, since the cumulative function $CD(\beta)$ is an invertible function of β and is uniformly distributed, $CD(\beta)$ constitutes a pivot conditional on x. Reciprocally, whenever a pivot increases with β (for example a continuous statistic $T(\beta)$ with cumulative

[3]For continuous distributions, just note that $P_\beta[\beta \le CD^{-1}(\alpha)] = P_\beta\{CD(\beta) \le CD\big(CD^{-1}(\alpha)\big)\} = P_\beta\{CD(\beta) \le \alpha]\} = \alpha$

distribution function F that is independent of β and free of any nuisance parameter), $F\big(T(\beta)\big)$ is uniformly distributed and satisfies conditions for providing a confidence distribution. Let $T(\beta)$ be such a continuous real statistic increasing with β with a free of nuisance parameter distribution. A test of $H_0 : \beta \leq \beta_0$ is rejected when $T^{obs}(\beta_0)$ is large, with p-value $P_{\beta_0}[T(\beta_0) > T^{obs}(\beta_0)]$. Then,

$$P_{\beta_0}[T(\beta_0) > T^{obs}(\beta_0)] = 1 - F_{\beta_0}(T^{obs}(\beta_0)) = CD(\beta_0) \tag{3.2}$$

where $F_{\beta_0}(.)$ is the sampling distribution of $T(\beta_0)$ under $\beta = \beta_0$. Consequently, simulated sampling distributions and simulated realized p-values as presented previously yield a way to construct simulated confidence distributions.

The sampling distribution and the confidence distribution are fundamentally different theoretical notions. The sampling distribution is the probability distribution of $T(\beta)$ obtained by repeated samplings whereas the confidence distribution is an ex-post object that contains the confidence statements one can have on the value of β given $y, x, T^{obs}(\beta)$.

Randomized confidence distributions for discrete statistics. A last remark relates to discrete statistics. Confidence distributions based on discrete statistics cannot lead to a continuous uniform distribution. Approximations must be used. Schweder and Hjort (2002) proposed half correction. For discrete statistics, they used

$$CD(\beta_0) = P_{\beta_0}[T(\beta_0) > T^{obs}(\beta_0)] + \frac{1}{2} P_{\beta_0}[T(\beta_0) = T^{obs}(\beta_0)], \tag{3.3}$$

We rather use randomization as in section 2. The discrete statistic $T(\beta)$ is associated with an auxiliary one U_T, which is independently, uniformly and continuously distributed over $[0,1]$. Lexicographical order is used to order ties.

$$CD(\beta_0) = P_{\beta_0}[T(\beta_0) > T^{obs}(\beta_0)] + P[U_{T(\beta_0)} > U_{T^{obs}(\beta_0)}] P_{\beta_0}[T(\beta_0) = T^{obs}(\beta_0)]. \tag{3.4}$$

Simulated confidence distributions and illustration. Let us consider a simple example to illustrate those notions. In the model $y_i = \beta x_i + u_i,\ i = 1, \ldots, n, (u_i, x_i) \overset{iid}{\sim} \mathcal{N}(0, I_2)$, the Student sign-based statistic

$$SST(\beta) = \frac{\sum s(y_i - x_i \beta) x_i}{(\sum x_i^2)^{1/2}}$$

is a pivotal function and decreases with β. The simulated confidence distribution of β given the realization y, x is

$$\widehat{CD}(\beta_0) = 1 - \hat{F}_{\beta_0}(SST(\beta_0)), \tag{3.5}$$

with $\hat{F}_{\beta_0}$ a Monte Carlo estimate of the sampling distribution of SST under $H_0(\beta_0) : \beta = \beta_0$. Figure 1 presents a simulated confidence distribution cumulative function for β, given 200 realizations of (u_i, x_i) based on SST. The Monte Carlo estimate of $\hat{F}_{\beta_0}$ is obtained from 9999 replicates of SST under $H_0(\beta_0)$.Testing $H_0^* : \beta \leq .1$ at 10% can be done by reading $CD(.1)$ here .92. The test accepts H_0^*. Further, $(-\infty, .23]$ constitutes a one-sided confidence interval for β with level .95.
Realized p-value functions for discrete statistics. Another interesting object is the realized p-value

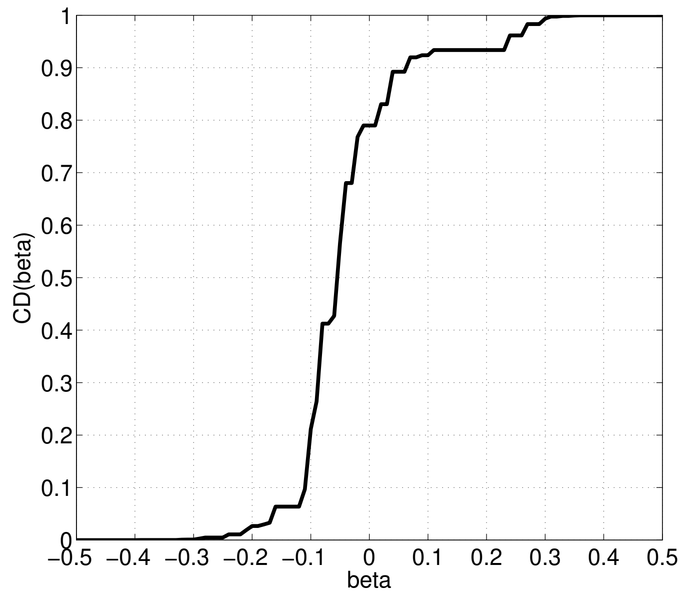


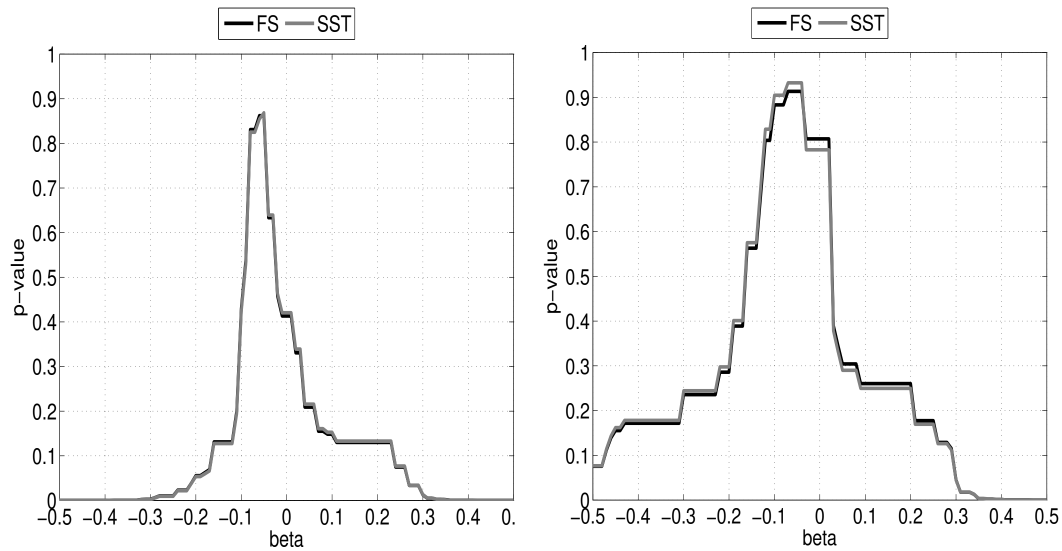
Figure 1. Simulated confidence distribution cumulative function based on SST.

function when testing point hypotheses $H_0(\beta_0)$. The latter is a simple transformation of the *CD* cumulative function. The simulated realized p-value is given by

$$\hat{p}_{SST}(\beta_0) = 2\min\{\widehat{CD}_{SST}(\beta_0), 1 - \widehat{CD}_{SST}(\beta_0)\}. \tag{3.6}$$

Consider now the statistic $SF = SST^2$. SF is a pivotal function but not a monotone function of β contrary to SST. An entire confidence distribution cannot be recovered from SF because of this lack of monotonicity. However, the p-value function can be constructed using equation (2.5). Figures 2 (a) and (b) compare p-value functions based on SST and SF. Inverting the p-value function allows one to recover half of the confidence distribution and consequently half of the inference results, *i.e.* the two-sided confidence intervals. For example, in Figure 2 (a), $[-.17, .24]$ constitutes a confidence interval with level 90% for both statistics. The p-value function provides then an interesting summary on the available inference. Especially, it gives the confidence degree one can have in the statement $\beta = \beta_0$. Finally, the p-value function has an important advantage over the confidence distribution: it is straightforwardly extendable to multidimensional parameters.

The spread of the p-value function is also related to the *model specification* and the *parameter identification*. When the p-value function is flat, one may expect the parameter to be badly identified either because there exists a set of observationally equivalent parameters, then, the p-values are high for a wide set of values; either because there does not exist any value satisfying the model and then the p-values are small everywhere. To illustrate that point, let us consider another example (example 2) where the first n_1 observations satisfy $y_i = \beta_1 x_i + u_i,\ i = 1, \ldots, n_1, (u_i, x_i) \overset{iid}{\sim} \mathcal{N}(0, I_2)$ and the n_2



(a) Example 1: well identified case (b) Example 2: misspecified case

Figure 2. Simulated p-value functions based on SST and SF

followings, $y_i = \beta_2 x_i + u_i,\ i = n_1+1,\ldots,n_1+n_2, (u_i, x_i) \overset{iid}{\sim} \mathscr{N}(0, I_2)$, with $\beta_1 = -.5$ and $\beta_2 = .5$. The model $y_i = \beta x_i + u_i,\ i = 1,\ldots,n_1+n_2$, is misspecified. In Figure 2 (b), we notice the spread of the p-value function based on *SF* is large: the set of observationally equivalent β is not reduced to a point.

3.2. Simultaneous and projection-based p-value functions in multivariate regression

If $p \geq 2$, the confidence distribution notion is not defined anymore. However, simulated realized p-values for testing $H_0(\beta_0)$ can easily be constructed from the *SF* statistic and more generally from any sign-based statistic which satisfies equation (2.4). Simulated p-values lead to a mapping for which we have a 3-dimensional representation for $p = 2$. Consider the model: $y_i = \beta^1 x_{1i} + \beta^2 x_{2i} + u_i,\ i = 1,\ldots,n, (u_i, x_{1i}, x_{2i}) \overset{iid}{\sim} \mathscr{N}(0, I_3)$, $\beta = (\beta^1, \beta^2) = (0,0)'$, $y = (y_1,\ldots,y_n)'$, $u = (u_1,\ldots,u_n)'$, $x_1 = (x_{11},\ldots,x_{1n})'$, $x_2 = (x_{21},\ldots,x_{2n})'$ and $X = (x_1, x_2)$. Let $D_S\big(\beta, (X'X)^{-1}\big) = s'(y - X\beta)X(X'X)^{-1}X's(y - X\beta)$. In Figure 3, we compute the simulated p-value function $\tilde{p}_N^{D_S}(\beta_0)$ for testing $H_0(\beta_0)$ on a grid of values of β_0, using N replicates of the sign vector. $\tilde{p}_N^{D_S}(\beta_0)$ allows one to construct simultaneous confidence sets for $\beta = (\beta^1, \beta^2)$ with any level. By construction, the confidence region $C_{1-\alpha}(\beta)$ defined as

$$C_{1-\alpha}(\beta) = \{\beta | \tilde{p}_N^{D_S}(\beta_0) \geq \alpha\}, \tag{3.7}$$

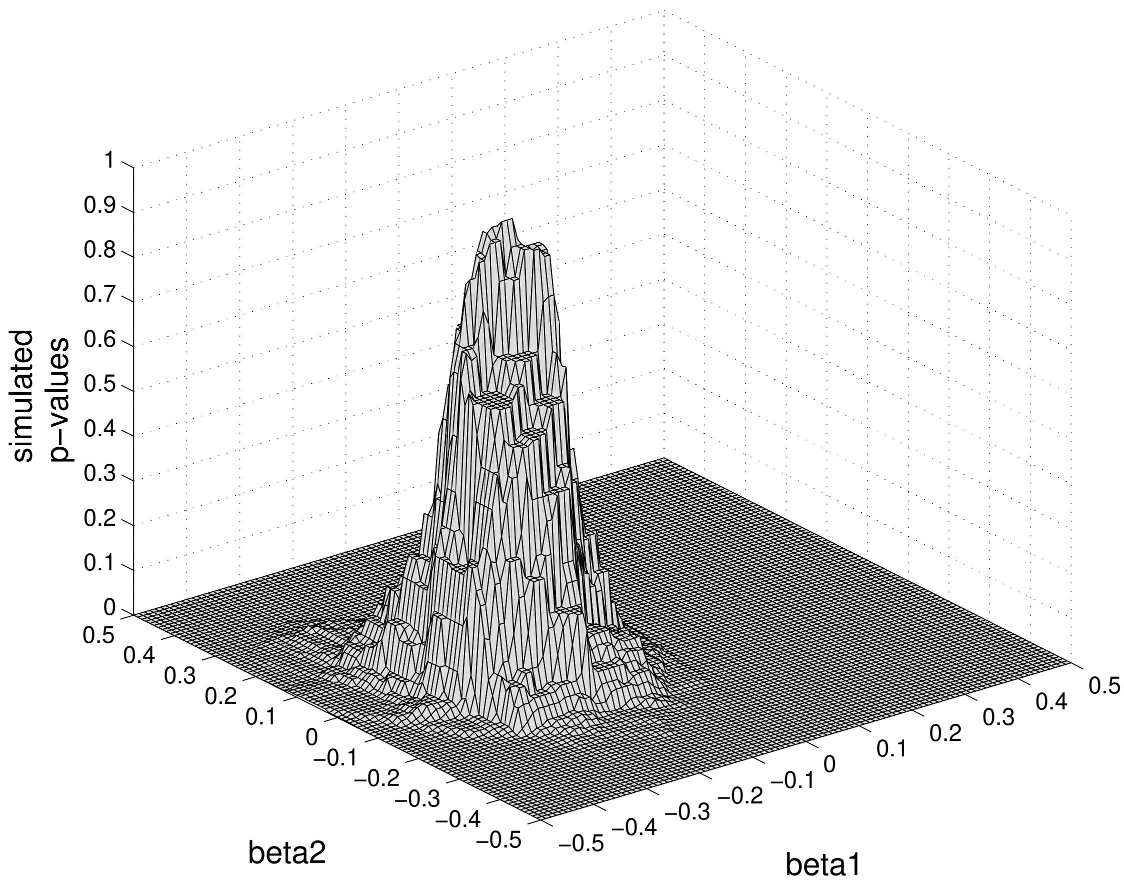


Figure 3. Simulated p-value functions based on SF ($n = 200$, $N = 9999$).

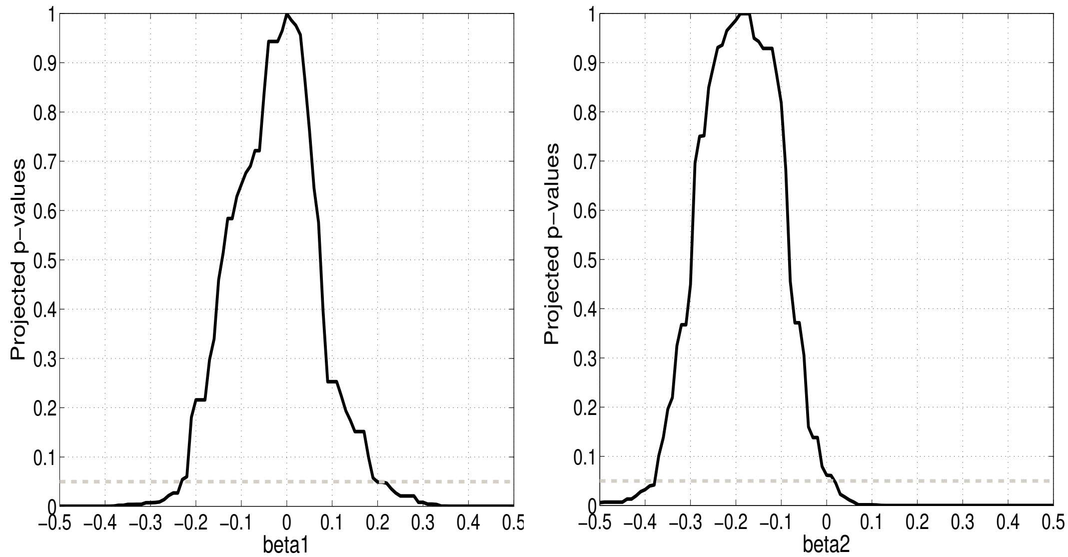
has level $1-\alpha$ [see Dufour (2006)]. Hence, by construction, $C_{1-\alpha}(\beta)$ corresponds to the intersection of the horizontal plan at ordinate α with the envelope of $\tilde{p}_N^{D_S}(\beta_0)$.

For higher dimensions ($p > 2$), a complete graphical representation is not available anymore. However, one can consider projection-based realized p-value functions for each individual component of the parameter of interest in a similar way than projection-based confidence intervals. For this, we apply the general strategy of projection on the complete simultaneous p-value function. The projected-based realized p-value function for the component β^1 is given by:

$$\text{Proj.}\tilde{p}_N^{\beta^1}(\beta_0^1) = \max_{\beta_0^2 \in \mathbb{R}} \tilde{p}_N^{D_S}[(\beta_0^1, \beta_0^2)]. \tag{3.8}$$

Figure 4 presents projection-based confidence intervals for the individual parameters of the previous

Figure 4. Projection-based p-values.



(a) Projection-based $p$values for β^1

(b) Projection-based $p$values for β^2

2-dimensional example. $[-.22, .21]$ is a 95% (conservative) confidence interval for β^1. $[-.38, .02]$ is a 95% (conservative) confidence interval for β^2. Testing $\beta^1 = 0$ is accepted at 5% with p-value 1.0. Testing $\beta^2 = 0$ is accepted at 5% with p-value .06.

Controlled inference using simulated confidence distributions and realized p-values. Simulated confidence distribution and realized p-values are Monte Carlo-based tools. Hence derived tests control the nominal size only for α's such that $\alpha(N+1) \in \mathbb{N}$; see Dufour (2006):

$$P[\tilde{p}_N^{Ds}(\beta_0) \leq \alpha] = \alpha \quad \forall\alpha \text{ such that } \alpha(N+1) \in \mathbb{N}.$$

If $\alpha(N+1) \notin \mathbb{N}$, only bounds on the significance level are known, but they are very close to α when N is sufficiently large:

$$\frac{I(\alpha(N+1)-1)}{N+1} \leq P[\tilde{p}_N^{Ds}(\beta_0) \leq \alpha] < \alpha \quad \forall\alpha \text{ such that } \alpha(N+1) \notin \mathbb{N}.$$

Contrary to tests, simulated confidence distributions and realized p-values are not evaluated at a given significance level α but rather on a range of significance levels $(\alpha_1, \ldots, \alpha_A)$. Hence, one must choose carefully N the number of replicates in order to control the significance level for all the α_i's, *i.e.* choose N sufficiently large to have $(N+1)\alpha_i \in \mathbb{N}, \quad \forall\alpha_i \in (\alpha_1, \ldots, \alpha_A)$. In the previous illustrations, $N = 9999$ which insures that the significance levels are controlled at .0001.

4. Sign-based estimators

Sign-based estimators complete the above system of inference. Intuition suggests to consider values with the highest confidence degree, *i.e*, with the highest p-values. Estimators obtained by that sort of test inversion constitute multidimensional extensions of the Hodges-Lehmann principle.

4.1. Sign-based estimators as maximizers of a p-value function

Hodges and Lehmann (1963) presented a general principle to derive estimators by test inversion; see also Johnson, Kotz and Read (1983). Suppose $\mu \in \mathbb{R}$ and $T(\mu_0, W)$ is a statistic for testing $\mu = \mu_0$ against $\mu > \mu_0$ based on the observations W. Suppose further that $T(\mu, W)$ is nondecreasing in the scalar μ. Given a known central value of $T(\mu_0, W)$, say $m(\mu_0)$ [for example $E_W T(\mu_0, W)$], the test rejects $\mu = \mu_0$ whenever the observed T is larger than, say, $m(\mu_0)$. If that is the case, one is inclined to prefer higher values of μ. The reverse holds when testing the opposite. If $m(\mu_0)$ does not depend on μ_0 [$m(\mu_0) = m_0$], an intuitive estimator of μ (if it exists) is given by μ^* such that $T(\mu^*, W)$ equals m_0 (or is very close to m_0). μ^* may be seen as the value of μ which is most supported by the observations.

This principle can be directly extended to multidimensional parameter setups through p-value functions. Let $\beta \in \mathbb{R}^p$. Consider testing $H_0(\beta_0) : \beta = \beta_0$ versus $H_1(\beta_0) : \beta = \beta_1$ with the positive statistic T. A test based on T rejects $H_0(\beta_0)$ when $T(\beta_0)$ is larger than a certain critical value that depends on the test level. The estimator of β is chosen as the value of β least rejected when the level α of the test increases. This corresponds to the highest p-value. If the associated p-value for $H_0(\beta_0)$ is $\mathrm{p}(\beta_0) = G\big(D_S(\beta_0)|\beta_0\big)$, where $G(x|\beta_0)$ is the survival function of $D_S(\beta_0)$, *i.e.* $G(x|\beta_0) = \mathrm{P}[D_S(\beta_0) > x]$, the set

$$M1 = \underset{\beta \in \mathbb{R}^p}{\arg\max}\ \mathrm{p}(\beta) \tag{4.1}$$

constitutes a set of Hodges-Lehmann-type estimators. HL-type estimators maximize the p-value function. There may not be a unique maximizer. In that case, any maximizer is consistent with the data.

4.2. Sign-based estimators as solutions of a nonlinear generalized least-squares problem

When the distribution of $T(\beta_0)$ and the corresponding p-value function do not depend on the tested value β_0, maximizing the p-value is equivalent to minimizing the statistic $T(\beta_0)$. This point is stated in the following proposition. Let us denote $\bar{F}(x|\beta_0)$ the distribution of $T(\beta_0)$ when $\beta = \beta_0$ and assume this distribution is invariant to β (Assumption 4.1).

Assumption 4.1 INVARIANCE OF THE DISTRIBUTION FUNCTION.

$$\bar{F}(x|\beta_0) = \bar{F}(x) \quad \forall x \in \mathbb{R}^+,\ \forall \beta_0 \in \mathbb{R}^p.$$

Let us define

$$M_1 = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmax}}\ \mathrm{p}(\beta). \tag{4.2}$$

$$M_2 = \underset{\beta \in \mathbb{R}^p}{\arg\min}\ T(\beta). \tag{4.3}$$

Then, the following proposition holds.

Proposition 4.1 *If Assumption 4.1 holds, then $M1 = M2$ with probability one.*

Maximizing $\mathrm{p}(\beta)$ is equivalent (in probability) to minimizing $T(\beta)$ if Assumption 4.1 holds. Under the mediangale Assumption 2.1, any sign-based statistic D_S does satisfy Assumption 4.1. Consequently,

$$\hat{\beta}_n(\Omega_n) \in \underset{\beta \in \mathbb{R}^p}{\arg\min}\ s'(Y - X\beta)X\Omega_n\big(s(Y - X\beta), X\big)X's(Y - X\beta) = M_2(Y, X, D_S^{\Omega_n}) \tag{4.4}$$

equals (with probability one) a Hodges-Lehmann estimator based on $D_S(\Omega_n, \beta)$. Since $D_S(\Omega_n, \beta)$ is non-negative, problem (4.4) always possesses at least one solution. As signs can only take 3 values, for fixed n, the quadratic function can take a finite number of values, which entails the existence of the minimum. If the solution is not unique, one may add a choice criterion. For example, one can choose the smallest solution in terms of a norm or use a randomization. Under conditions of point identification, any solution of (4.4) is a consistent estimator.

In models with sets of observationally equivalent values of β, any inference approach relying on the consistency of a point estimator (which assumes point identification), gives misleading results whereas a whole estimator set remains informative. The approach of Chernozhukov, Hong and Tamer (2007) can be applied here. Let us remind that the Monte Carlo sign-based inference method [Coudin and Dufour (2009)] does not rely on identification conditions and leads to valid results in any case.

The sign-based estimators studied by Boldin et al. (1997) are solutions of

$$\hat{\beta}_n(I_p) \in\ \arg\min_{\beta \in \mathbb{R}^p} s'(Y - X\beta)XX's(Y - X\beta) = \arg\min_{\beta \in \mathbb{R}} SB(\beta), \tag{4.5}$$

and

$$\hat{\beta}_n((X'X)^{-1}) \in\ \arg\min_{\beta \in\ \mathbb{R}^p} s'(Y - X\beta)X(X'X)^{-1}X's(Y - X\beta) = \arg\min_{\beta \in \mathbb{R}} SF(\beta). \tag{4.6}$$

For heteroskedastic independent disturbances, we introduce weighted versions of sign-based estimators that can be more efficient than the basic ones defined in (4.5) or (4.6). Weighted sign-based estimators are sign-based analogues to weighted LAD estimator [see Zhao (2001)]. The weighted LAD estimator is given by

$$\beta_n^{WLAD} =\ \underset{\beta \in \mathbf{R}^p}{\operatorname{argmin}} \sum_i d_i |y_i - x_i'\beta|. \tag{4.7}$$

The weighted sign-based estimators are solutions of

$$\hat{\beta}_n^{DX} \in \underset{\beta\in\mathbf{R}^p}{\operatorname{argmin}}\, s'(Y-X\beta)\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'D's(Y-X\beta) \tag{4.8}$$

where $\tilde{X}=\operatorname{diag}(d_1,\ldots,d_n)X$ and $(d_i), i=1,\ldots,n\in\mathbb{R}^{+*}$. Weighted sign-based estimators that involve optimal estimating functions in the sense of Godambe (2001) are solutions of

$$\hat{\beta}_n^{DX^*} \in \underset{\beta\in\mathbf{R}^p}{\operatorname{argmin}}\, s'(Y-X\beta)X^*(X^{*'}X^*)^{-1}X^{*'}D's(Y-X\beta) \tag{4.9}$$

where $\tilde{X}=\operatorname{diag}\big(f_1(0|X),\ldots,f_n(0|X)\big)X$ and $f_t(0|X), t=1,\ldots,n$, are the conditional disturbance densities evaluated at zero. The inherent problem of such a class of estimators is to provide good approximations of $f_i(0|X)$'s. Densities of normal distributions can be used.

4.3. Sign-based estimators as GMM estimators

Sign-based estimators have been interpreted in the literature as GMM estimators exploiting the orthogonality condition between the signs and the explanatory variables [see Honore and Hu (2004)]. In our opinion, a strictly GMM interpretation hides the link with the testing theory. That is the reason why we first introduced sign-based estimators as Hodges-Lehmann estimators. The quadratic form (4.4) refers to quite unusual moment conditions. The sign transformation evacuates the unknown parameters that affect the error distribution. It validates nonparametric finite-sample-based inference when mediangale Assumption holds. However, in settings where only the sign-moment condition 2.2 is satisfied, the GMM interpretation of sign-based estimators still applies and entails useful extensions.

For autocorrelated disturbances, an estimator based on a HAC sign-based statistic $D_S(\beta,\hat{J}_n^{-1})$ can be used:

$$\hat{\beta}_n(\hat{J}_n^{-1}) \in \arg\min_{\beta\in\mathbf{R}^p} s'(Y-X\beta)X[\hat{J}_n(s(Y-X\beta),X)]^{-1}X's(Y-X\beta), \tag{4.10}$$

where $\hat{J}_n(s(Y-X\beta),X)$ accounts for the dependence among the signs and the explanatory variables. β appears twice, first in the constrained signs, second in the weight matrix. In practice, optimizing (4.10) requires one to invert a new matrix $\hat{J}_n$ for each value of β whereas problem (4.6) only requires one inversion of $X'X$. In practice, this numerical problem may quickly become cumbersome similarly to continuously updating GMM. We advocate to use a two-step method: first, solve (4.6) and obtain $\hat{\beta}_n((X'X)^{-1})$; compute then $\hat{J}_n^{-1}\big(s(Y-X\hat{\beta}_n((X'X)^{-1})),X\big)$ and finally solve,

$$\hat{\beta}_n^{2S}(\hat{J}_n^{-1}) \in \arg\min_{\beta\in\mathbf{R}^p} s'(Y-X\beta)X[\hat{J}_n(s(Y-X\hat{\beta}_n),X)]^{-1}X's(Y-X\beta). \tag{4.11}$$

The 2-step estimator is not a Hodges-Lehmann estimator anymore. However, it is still consistent and share some interesting finite-sample properties with classical sign-based estimators. The properties of sign-based estimators are studied in the next section.

5. Finite-sample properties of sign-based estimators

In this section, finite-sample properties of sign-based estimators are studied. Sign-based estimators share invariance properties with the LAD estimator and are median-unbiased if the disturbance distribution is symmetric and some additional assumptions on the form of the solution are satisfied. The topology of the argmin set of the optimization problem 4.4 does not possess a simple structure. In some cases it is reduced to a single point like the empirical median of $2p+1$ observations. In other cases, it is a set. More generally, the argmin set is a union of convex sets but it is not *a priori* either convex nor connected. To see that it is a union of convex sets just remark that the reciprocal image of n fixed signs is convex.

Sign-based estimators share some attractive equivariance properties with LAD and quantile estimators [see Koenker and Bassett (1978)]. It is straightforward to see that the following proposition holds.

Proposition 5.1 INVARIANCE. *Let $M(y,X)$ be the set of the solutions of the minimization problem* (4.4). *If $\hat{\beta}(y,X) \in M(y,X)$, then the following properties hold:*

$$\begin{aligned}
\lambda\hat{\beta}(y,X) &\in M(\lambda y, X), \quad \forall \lambda \in \mathbb{R}, && (5.1)\\
\hat{\beta}(y,X)+\gamma &\in M(y+X\gamma, X), \quad \forall \gamma \in \mathbb{R}^p, && (5.2)\\
A^{-1}\hat{\beta}(y,X) &\in M(y, XA), \quad \textit{for any nonsingular } k\times k \textit{ matrix } A. && (5.3)
\end{aligned}$$

Further, if $\hat{\beta}(y,X)$ is a uniquely determined solution of (4.4), *then*

$$\begin{aligned}
\hat{\beta}(\lambda y,X) &= \lambda\hat{\beta}(y,X), \quad \forall \lambda \in \mathbb{R}, && (5.4)\\
\hat{\beta}(y+X\gamma,X) &= \hat{\beta}(y,X)+\gamma, \quad \forall \gamma \in \mathbb{R}^p, && (5.5)\\
\hat{\beta}(y,XA) &= A^{-1}\hat{\beta}(y,X), \quad \textit{for any nonsingular } k\times k \textit{ matrix } A. && (5.6)
\end{aligned}$$

To prove this property, it is sufficient to write down the different optimization problems. (5.1) and (5.4) state a form of scale invariance: if y is rescaled by a certain factor, $\hat{\beta}$, rescaled by the same one is solution of the transformed problem. (5.2) and (5.5) represent location invariance, while (5.3) and (5.6) show the behavior of the estimator changes states a reparameterization of the design matrix. In all cases, parameter estimates change in the same way as theoretical parameters.

If the disturbance distribution is assumed to be symmetric and the optimization problems to have a unique solution then sign-estimators are median unbiased.

Proposition 5.2 MEDIAN UNBIASEDNESS. *If $u \sim -u$ and the sign-based estimator $\hat{\beta}(y,X)$ is a uniquely determined solution of the minimization problem*(4.4)*, then $\hat{\beta}$ is median unbiased,* i.e.

$$\mathrm{Med}(\hat{\beta}-\bar{\beta}) = 0$$

where $\bar{\beta}$ represents the "true value" of β.

6. Asymptotic properties

We demonstrate consistency of the proposed sign-based estimators when the parameter is identified under weaker assumptions than the LAD estimator, which validates the use of sign-based estimators even in settings when the LAD estimator fails to converge. Finally, sign-based estimators are asymptotically normal. For reviews of the asymptotic distributional theory of LAD estimators, the reader may consult Bassett and Koenker (1978), Knight (1989), Phillips (1991), Pollard (1991), Portnoy (1991), Weiss (1991), Fitzenberger (1997), Knight (1998), El Bantli and Hallin (1999), and Koenker (2005).

6.1. Identification and consistency

We show that the sign-based estimators (4.4) and (4.11) are consistent under the following set of assumptions. In the sequel, we denote by $\bar{\beta}$ the "true value" of β, and by β_0 any hypothesized value.

Assumption 6.1 MIXING. *$\{W_t = (y_t, x_t')\}_{t=1,2,\ldots}$ is α-mixing of size $-r/(r-1)$ with $r > 1$.*

Assumption 6.2 BOUNDEDNESS. *$x_t = (x_{1t}, \ldots, x_{pt})'$ and $\mathsf{E}|x_{ht}|^{r+1} < \Delta < \infty$, $h = 1, \ldots, p$, $t = 1, \ldots, n$, $\forall n \in \mathbb{N}$.*

Assumption 6.3 COMPACTNESS. *$\bar{\beta} \in \mathrm{Int}(\Theta)$, where Θ is a compact subset of $\mathbb{R}^p$.*

Assumption 6.4 REGULARITY OF THE DENSITY.

1. *There are positive constants f_L and p_1 such that, for all $n \in \mathbb{N}$,*

$$\mathsf{P}[f_t(0\,|\,X) > f_L] > p_1,\ t = 1, \ldots, n,\ a.s.$$

2. *$f_t(\cdot\,|\,X)$ is continuous, for all $n \in \mathbb{N}$ for all t, a.s.*

Assumption 6.5 POINT IDENTIFICATION CONDITION. *$\forall \delta > 0, \exists \tau > 0$ such that*

$$\liminf_{n\to\infty} \frac{1}{n} \sum_t \mathsf{P}[|x_t'\delta| > \tau\,|\,f_t(0\,|\,x_1, \ldots, x_n) > f_L] > 0.$$

Assumption 6.6 UNIFORMLY POSITIVE DEFINITE WEIGHT MATRIX. *$\bar{\Omega}_n(\beta)$ is symmetric positive definite for all β in Θ.*

Assumption 6.7 LOCALLY POSITIVE DEFINITE WEIGHT MATRIX. *$\bar{\Omega}_n(\beta)$ is symmetric positive definite for all β in a neighborhood of $\bar{\beta}$.*

Then, we can state the consistency theorem. The assumptions are interpreted just after.

Theorem 6.1 CONSISTENCY. *Under model* (2.1) *with the assumptions 2.2 and 6.1-6.6, any sign-based estimator of the type,*

$$\hat{\beta}_n \in \underset{\beta_0 \in \Theta}{\operatorname{argmin}}\, s(y - X\beta_0)' X \Omega_n \left[s(y - X\beta_0), X \right] X' s(y - X\beta_0) \tag{6.1}$$

or

$$\hat{\beta}_n^{2S} \in \underset{\beta_0 \in \Theta}{\operatorname{argmin}}\, s(y - X\beta_0)' X \hat{\Omega}_n \left[s(y - X\hat{\beta}), X \right] X' s(y - X\beta_0), \tag{6.2}$$

where $\hat{\beta}$ *stands for any (first step) consistent estimator of* $\bar{\beta}$*, is consistent.* $\hat{\beta}_n^{2S}$ *defined in equation* (6.2) *is also consistent if Assumption 6.6 is replaced by Assumption 6.7.*

It will useful to discuss Assumptions 6.1 - 6.7 and compare them with the ones required for LAD and quantile estimator consistency. On considering the special case where $X\Omega_n\left[s(y - X\beta_0), X\right]X' = I_n$ the identity matrix, the estimators in (6.1) - (6.2) coincide with the "quantile regression estimator" (with $\theta = 1/2$) studied by Fitzenberger (1997, Theorem 2.2). However, allowing for a weighting matrix different the identity matrix – as we do here – turns out to be important from the viewpoint of efficiency. *Stricto sensu*, the sign-based estimators in (6.1) - (6.2) and Fitzenberger (1997, Theorem 2.2) are not LAD estimators, because the size of residuals (through absolute values) do not appear in the objective function. This feature is crucial for relaxing assumptions on moments. The disturbances indeed appear in the objective function only through their sign transforms which possess finite moments at all orders. Consequently, no additional restriction need be imposed on the disturbance process (in addition to regularity conditions on the density). Only assumptions on the moments of x_t are used (see Assumption 6.2). There is very little work on LAD estimators properties with infinite variance errors; see Knight (1989) and Phillips (1991) who derive LAD asymptotic properties for an autoregressive model with infinite variance errors, which are in the domain of attraction of a stable law.

Assumption 6.1 on mixing is needed to apply a generic weak law of large numbers; see Andrews (1987) and White (2001). It was used by Fitzenberger (1997) with stationary linearly dependent processes. It covers, among other processes, stationary ARMA disturbances with continuously distributed innovations. Identification is provided by Assumptions 6.4 and 6.5. Assumption 6.5 is similar to Condition ID in Weiss (1991). Assumption 6.4 is usual in LAD estimator asymptotics.[4] It is analogous to Fitzenberger's (1997) conditions (ii.b) - (ii.c) and Weiss's (1991) CD condition. It implies that there is enough variation around zero to identify the median. It restricts the setup for some "bounded" heteroskedasticity in the disturbance process but not in the usual (variance-based) way. It is related to *diffusivity* $\left(2f(0)\right)^{-1}$, a dispersion measure adapted to median-unbiased estimators. Diffusivity indicates the vertical spread of a density rather than its horizontal spread, and appears in Cramér-Rao-type efficiency bounds for median-unbiased estimators; see Sung, Stangenhaus and David (1990) and So (1994). Assumption 6.6 entails that the weight matrix Ω_n is everywhere invertible, while Assumption 6.7 only requires local invertibility.

[4] Assumption 6.4 can be slightly relaxed covering error terms with mass point if the objective function involves randomized signs instead of usual signs.

6.2. Asymptotic normality

Sign-based estimators are asymptotically normal. Sign-based estimators are well adapted to deal with heavy-tailed disturbances that may not possess finite variances. The assumptions we consider are the following ones.

Assumption 6.8 UNIFORMLY BOUNDED DENSITIES. *$\exists f_U < +\infty$ such that ,$\forall n \in \mathbb{N}, \forall \lambda \in \mathbb{R}$,*

$$\sup_{\{t \in (1,\dots,n)\}} |f_t(\lambda \,|\, x_1, \dots, x_n)| < f_U, \ a.s.$$

Under the conditions 2.2, 6.1, 6.2 and 6.8, we can define $L(\beta)$, the derivative of the limiting objective function at β:

$$L(\beta) = \lim_{n\to\infty} \frac{1}{n} \sum_t \mathsf{E}\big[x_t x_t' f_t\big(x_t'(\beta - \bar{\beta}) \,|\, x_1, \dots, x_n\big)\big] = \lim_{n\to\infty} L_n(\beta). \tag{6.3}$$

where

$$L_n(\beta) = \frac{1}{n} \sum_t \mathsf{E}\big[x_t x_t' f_t\big(x_t'(\beta - \bar{\beta}) \,|\, x_1, \dots, x_n\big)\big]. \tag{6.4}$$

The other assumptions are fairly standard conditions to prove asymptotic normality.

Assumption 6.9 MIXING WITH $r > 2$. *The process $\{W_t = (y_t, x_t') : t = 1, 2, \dots\}$ is α-mixing of size $-r/(r-2)$ with $r > 2$.*

Assumption 6.10 DEFINITE POSITIVENESS OF L_n. *$L_n(\bar{\beta})$ is positive definite uniformly in n.*

Assumption 6.11 DEFINITE POSITIVENESS OF J_n. *$J_n = \mathsf{E}\big[\frac{1}{n}\sum_{t,s}^n s(u_t) x_t x_s' s(u_s)\big]$ is positive definite uniformly in n and converges to a definite positive symmetric matrix J as $n \to \infty$.*

Then, we have the following result.

Theorem 6.2 ASYMPTOTIC NORMALITY. *Under the assumptions* (2.2)*, 6.1 to 6.6, and 6.9 to 6.11, we have:*

$$S_n^{-1/2} \sqrt{n} \big[\hat{\beta}_n - \bar{\beta}\big] \xrightarrow{d} \mathrm{N}(0, I_p) \tag{6.5}$$

where $\hat{\beta}_n(\Omega_n)$ is any estimator which minimizes $D_S[\beta_0, \bar{\Omega}_n(\beta_0)]$ in (2.4),

$$S_n = [L_n(\bar{\beta})\Omega_n L_n(\bar{\beta})]^{-1} L_n(\bar{\beta})\Omega_n J_n \Omega_n L_n(\bar{\beta})[L_n(\bar{\beta})\Omega_n L_n(\bar{\beta})]^{-1}$$

and

$$L_n(\bar{\beta}) = \frac{1}{n} \sum_t \mathsf{E}\big[x_t x_t' f_t\big(0 \,|\, x_1, \dots, x_n\big)\big]. \tag{6.6}$$

When $\bar{\Omega}_n(\beta_0) = \hat{J}_n(\beta_0)^{-1}$ and $\hat{J}_n(\beta_0) = \frac{1}{n}\sum_{t,s}^n s(y_t - x_t'\beta_0) x_t x_s' s(y_s - x_s'\beta_0)$, we get:

$$[L_n(\bar{\beta}) \hat{J}_n^{-1} L_n(\bar{\beta})]^{-1/2} \sqrt{n} \big[\hat{\beta}_n(\hat{J}_n^{-1}) - \bar{\beta}\big] \xrightarrow{d} \mathrm{N}\big[0, I_p\big]\,. \tag{6.7}$$

This corresponds to the use of optimal instruments and quasi-efficient estimation. $\hat{\beta}_n(\hat{J}_n^{-1})$ has the same asymptotic covariance matrix as the LAD estimator. Thus, performance differences between the two estimators correspond to finite-sample features. This result contradicts the generally accepted idea that sign procedures involve a heavy loss of information. There is no loss induced by the use of signs instead of absolute values.

Note again that we do not require that the disturbance process variance be finite. We only assume that the second-order moments of X are finite and the mixing property of $\{W_t,\ t = 1,\ldots\}$ holds. This differs from usual assumptions for LAD asymptotic normality.[5] This difference comes from the fact that absolute values of the disturbance process are replaced in the objective function by their signs. Since signs possess finite moments at any order, one sees easily that a CLT can be applied without any further restriction. Consequently, asymptotic normality, such as consistency, holds for heavy-tailed disturbances that may not possess finite variance. This is an important theoretical advantage of sign-based rather than absolute value-based estimators and, *a fortiori*, rather than least-squares estimators. Estimators, for which asymptotic normality holds on bounded asymptotic variance assumption (for example OLS) are not accurate in heavy-tail settings because the variance is not a measure of dispersion adapted to those settings. Estimators, for which the asymptotic behavior relies on other measures of dispersion, like the diffusivity, help one out of trouble.

The form of the asymptotic covariance matrix simplifies under stronger assumptions. When the signs are mutually independent conditional on X [mediangale Assumption 2.1], both $\hat{\beta}_n((X'X)^{-1})$ and $\beta(\hat{J}_n^{-1})$ are asymptotically normal with variance

$$S_n = [L_n(\bar{\beta})]^{-1}\mathsf{E}\left[(1/n)\sum_{t=1}^{n} x_t x_t'\right][L_n(\bar{\beta})]^{-1}.$$

If u is an *i.i.d.* process and is independent of X, then $f_t(0) = f(0)$, and

$$S_n = \frac{1}{4f(0)^2}\mathsf{E}(x_t x_t')^{-1}. \tag{6.8}$$

In the general case, $f_t(0)$ is a nuisance parameter even if condition 6.8 implies that it can be bounded.

All the features known about the LAD estimator asymptotic behavior apply also for the *SHAC* estimator; see Boldin et al. (1997). For example, asymptotic relative efficiency of the *SHAC* (and LAD) estimator with respect to the OLS estimator is $2/\pi$ if the errors are normally distributed $N(0,\sigma^2)$, but *SHAC* (such as LAD) estimator can have arbitrarily large ARE with respect to OLS when the disturbance generating process is contaminated by outliers.

6.3. Asymptotic or projection-based confidence sets?

In section 4, we introduced sign-based estimators as Hodges-Lehmann estimators associated with sign-based statistics. By linking them with GMM settings, we then derived asymptotic normality. We stressed that sign-based estimator asymptotic normality holds under weaker assumptions

[5] See Fitzenberger (1997) for the derivation of the LAD asymptotics in a similar setup and Bassett-Koenker(1978) or Weiss (1991) for a derivation of the LAD asymptotics under sign independence.

than the ones needed for the LAD estimator. Therefore, sign-based estimator asymptotic normality enables one to construct asymptotic tests and confidence intervals. Thus, we have two ways of making inference with signs: we can use the Monte Carlo (finite-sample) based method described in Coudin and Dufour (2009) - see section 2.2 - and the classical asymptotic method. Let us list here the main differences between them. Monte Carlo inference relies on the pivotality of the sign-based statistic. The derived tests are valid (with controlled level) for any sample size if the mediangale Assumption 2.1 holds. When only the sign moment condition 2.2 holds, the Monte Carlo inference remains asymptotically valid. Asymptotic test levels are controlled. Besides, in simulations, the Monte Carlo inference method appears to perform better in small samples than classical asymptotic methods, even if its use is only asymptotically justified [see Coudin and Dufour (2009)]. Nevertheless, that method has an important drawback: its computational complexity. On the contrary, classical asymptotic methods which yield tests with controlled asymptotic level under the sign moment condition 2.2 may be less time consuming. The choice between both is mainly a question of computational capacity. We point out that classical asymptotic inference greatly relies on the way the asymptotic covariance matrix, that depends on unknown parameters (densities at zero), is treated. If the asymptotic covariance matrix is estimated thanks to a simulation-based method (such as the bootstrap) then the time argument does not hold anymore. Both methods would be of the same order of computational complexity.

7. Simulation study

In this section, we compare the performance of the sign-based estimators with the OLS and LAD estimators in terms of asymptotic bias and RMSE.

7.1. Simulation setup

We use estimators derived from the sign-based statistics $D_S\big(\beta,(X'X)^{-1}\big)$ and $D_S(\beta,\hat{J}_n^{-1})$ when a correction is needed for linear serial dependence (*SHAC* estimator). Minimizations are solved by simulated annealing. We consider a set of general DGP's to illustrate different classical problems one may encounter in practice. We use the following linear regression model:

$$y_t = x_t'\beta + u_t \tag{7.1}$$

where $x_t = (1, x_{2,t}, x_{3,t})'$ and β are 3×1 vectors. We denote the sample size n. Monte Carlo studies are based on S generated random samples. Table 1 presents the cases considered.

In a first group of examples (A1-A4), we consider classical independent cases with bounded heterogeneity. In a second one (B5-B8), we look at processes involving large heteroskedasticity so that some of the estimators we consider may not be asymptotically normal nor even consistent. Finally, the third group (C9-C11) is dedicated to autocorrelated disturbances. We wonder whether the two-step *SHAC* sign-based estimator performs better in small samples than the non-corrected one.

To sum up, cases A1 and A2 present *i.i.d.* normal observations without and with conditional heteroskedasticity. Case A3 involves a sort of weak nonlinear dependence in the error term. Case A4

Table 1. Simulated models.

A1:	Normal HOM errors	$(x_{2,t},x_{3,t},u_t)' \overset{i.i.d}{\sim} \mathcal{N}(0,I_3),\ t=1,\ldots,n$
A2:	Normal HET errors	$(x_{2,t},x_{3,t},\tilde{u}_t)' \overset{i.i.d}{\sim} \mathcal{N}(0,I_3)$, $u_t = \min\{3, \max[0.21,\lvert x_{2,t}\rvert]\} \times \tilde{u}_t,\ t=1,\ldots,n$
A3:	Dep.-HET $\rho_x = .5$:	$x_{j,t} = \rho_x x_{j,t-1} + \nu_t^j,\ j=1,2,$ $u_t = \min\{3,\max[0.21,\lvert x_{2,t}\rvert]\} \times \nu_t^u,$ $(\nu_t^2,\nu_t^3,\nu_t^u)' \overset{i.i.d}{\sim} \mathcal{N}(0,I_3),\ t=2,\ldots,n$ ν_1^2 and ν_1^3 chosen to insure stationarity.
A4:	Unbalanced design matrix	$x_{2,t} \sim \mathscr{B}(1,0.3),\ x_{3,t} \overset{i.i.d.}{\sim} \mathcal{N}(0,.01^2),$ $u_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1),\ x_t,u_t$ independent, $t=1,\ldots,n$.
B5:	Cauchy errors	$(x_{2,t},x_{3,t})' \sim \mathcal{N}(0,I_2),$ $u_t \overset{i.i.d.}{\sim} \mathscr{C}, x_t,u_t,$ independent, $t=1,\ldots,n$.
B6:	Stochastic volatility	$(x_{2,t},x_{3,t})' \overset{i.i.d.}{\sim} \mathcal{N}(0,I_2), u_t = exp(w_t/2)\varepsilon_t$ with $w_t = 0.5w_{t-1} + v_t$, where $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1),\ v_t \overset{i.i.d.}{\sim} \chi_2(3),$ $x_t,u_t,$ independent, $t=1,\ldots,n$.
B7:	Nonstationary GARCH(1,1)	$(x_{2,t},x_{3,t},\varepsilon_t)' \overset{i.i.d.}{\sim} \mathcal{N}(0,I_3),\ t=1,\ldots,n,$ $u_t = \sigma_t\varepsilon_t,\ \sigma_t^2 = 0.8u_{t-1}^2 + 0.8\sigma_{t-1}^2.$
B8:	Exponential error variance	$(x_{2,t},x_{3,t},\varepsilon_t)' \overset{i.i.d.}{\sim} \mathcal{N}(0,I_3),\ u_t = exp(.2t)\varepsilon_t.$
C9:	AR(1)-HOM $\rho_u = .5$	$(x_{2,t},x_{3,t},\nu_t^u)' \sim \mathcal{N}(0,I_3),\ t=2,\ldots,n,$ $u_t = \rho_u u_{t-1} + \nu_t^u,$ $(x_{2,1},x_{3,1})' \sim \mathcal{N}(0,I_2),\ \nu_1^u$ insures stationarity.
C10:	AR(1)-HET $\rho_u = .5,$ $\rho_x = .5$	$x_{j,t} = \rho_x x_{j,t-1} + \nu_t^j,\ j=1,2,$ $u_t = min\{3,\max[0.21,\lvert x_{2,t}\rvert]\} \times \tilde{u}_t,$ $\tilde{u}_t = \rho_u \tilde{u}_{t-1} + \nu_t^u,$ $(\nu_t^2,\nu_t^3,\nu_t^u)' \overset{i.i.d}{\sim} \mathcal{N}(0,I_3),\ t=2,\ldots,n$ ν_1^2,ν_1^3 and ν_1^u chosen to insure stationarity.
C11:	AR(1)-HOM $\rho_u = .9$	$(x_{2,t},x_{3,t},\nu_t^u)' \sim \mathcal{N}(0,I_3), t=2,\ldots,n,$ $u_t = \rho_u u_{t-1} + \nu_t^u,$ $(x_{2,1},x_{3,1})' \sim \mathcal{N}(0,I_2),\ \nu_1^u$ insures stationarity.

presents a very unbalanced scheme in the design matrix (a case when the LAD estimator is known to perform badly). Cases B5, B6, B7 and B8 are other cases of long tailed errors, heteroskedasticity and nonlinear dependence. Cases C9 to C11 illustrate different levels of autocorrelation in the error term with and without heteroskedasticity.

7.2. Bias and RMSE

We give biases and RMSE of each parameter of interest in Table 2 and we report a norm of these three values. $n = 50$ and $S = 1000$. These results are unconditional on X.

In classical cases (A1-A3), sign-based estimators have roughly the same behavior as the LAD estimator, in terms of bias and RMSE. OLS is optimal in case A1. However, there is no important efficiency loss or bias increase in using signs instead of LAD. Besides, if the LAD is not accurate in a particular setup (for example with highly unbalanced explanatory scheme, case A4), the sign-based estimators do not suffer from the same drawback. In case A4, the RMSE of the sign-based estimator is notably smaller than those of the OLS and the LAD estimates.

For setups with strong heteroskedasticity and nonstationary disturbances (B5-B8), we see that the sign-based estimators yield better results than both LAD and OLS estimators. Not far from the (optimal) LAD in case of Cauchy disturbances (B5), the signs estimators are the only estimators that stay reliable with nonstationary variance (B6-B8). No assumption on the moments of the error term is needed for sign-based estimators consistency. All that matters is the behavior of their signs.

When the error term is autocorrelated (C9-C11), results are mixed. When a moderate linear dependence is present in the data, sign-based estimators give good results (C9, C10). But when the linear dependence is stronger (C11), that is no longer true. The *SHAC* sign-based estimator does not give better results than the non-corrected one in these selected examples.

To conclude, sign-based estimators are robust estimators much less sensitive than the LAD estimator to various unbalanced schemes in the explanatory variables and to heteroskedasticity. They are particularly adequate when an amount an heteroskedasticity or nonlinear dependence is suspected in the error term, even if the error term fails to be stationary. Finally, the HAC correction does not seem to increase the performance of the estimator. Nevertheless, it does for tests. We show in Coudin and Dufour (2009) that using a HAC-corrected statistic allows for the asymptotic validity of the Monte Carlo inference method and improves the test performance in small samples.

8. Empirical applications

In this section, we go back to the two illustrations presented in Coudin and Dufour (2007, 2009) where sign-based tests were derived, with now estimation in mind. The first application is dedicated to estimate a drift on the Standard and Poor's Composite Price Index (S&P), 1928-1987. In the second one, we search a robust estimate of the rate of β-convergence between output levels across U.S. States during the 1880-1988 period using Barro and Sala-i-Martin (1991) data.

Table 2. Simulated bias and RMSE.

$n = 50$		OLS		LAD		SF		SHAC	
$S = 1000$		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
A1:	β_0	.003	.142	.002	.179	.002	.179	.004	.178
	β_1	.003	.149	.006	.184	.004	.182	.004	.182
	β_2	$-.002$	.149	$-.007$	.186	$-.006$	.185	$-.007$	.183
	$\lVert\beta\rVert$*	**.004**	**.254**	**.009**	**.316**	**.007**	**.315**	**.009**	**.313**
A2:	β_0	$-.003$	.136	.000	.090	$-.000$	.089	$-.000$	.089
	β_1	$-.0135$	.230	$-.006$	.218	$-.010$	.218	$-.010$	.218
	β_2	.002	.142	$-.001$	.095	$-.001$	.092	$-.001$	.092
	$\lVert\beta\rVert$	**.014**	**.303**	**.007**	**.254**	**.010**	**.253**	**.010**	**.253**
A3:	β_0	.022	.167	.018	.108	.025	.107	.023	.107
	β_1	-1.00	.228	.005	.215	.003	.214	.002	.215
	β_2	.001	.150	.005	.105	.007	.104	.007	.105
	$\lVert\beta\rVert$	**.022**	**.320**	**.019**	**.263**	**.026**	**.261**	**.024**	**.262**
A4:	β_0	$-.001$	.174	.007	.2102	.010	.2181	.008	.2171
	β_1	$-.016$	.313	$-.011$	.375	$-.021$	.396	$-.021$	.394
	β_2	$-.100$	14.6	.077	18.4	.014	7.41	.049	7.40
	$\lVert\beta\rVert$	**.101**	**14.6**	**.078**	**18.5**	**.027**	**7.42**	**.054**	**7.41**
B5:	β_0	16.0	505	.001	.251	.004	.248	.003	.248
	β_1	-3.31	119	.015	.264	.020	.265	.020	.265
	β_2	-2.191	630	.000	.256	.003	.258	.001	.258
	$\lVert\beta\rVert$	**26.0**	**817**	**.015**	**.445**	**.021**	**.445**	**.020**	**.445**
B6:	β_0	$-.908$	29.6	-1.02	27.4	.071	2.28	.083	2.28
	β_1	2.00	37.6	3.21	68.4	.058	2.38	.069	2.39
	β_2	1.64	59.3	2.59	91.8	$-.101$	2.30	$-.089$	2.29
	$\lVert\beta\rVert$	**2.73**	**76.2**	**4.25**	**118**	**.136**	**4.02**	**.139**	**4.02**
B7:	β_0	-127	3289	$-.010$	7.85	$-.008$	3.16	$-.028$	3.17
	β_1	-81.4	237	.130	11.2	$-.086$	3.80	$-.086$	3.823
	β_2	-31.0	1484	$-.314$	12.0	$-.021$	3.606	$-.009$	3.630
	$\lVert\beta\rVert$	**154**	**4312**	**.340**	**18.2**	**.089**	**6.12**	**.091**	**6.15**
B8:	β_0	$< -10^{10}$	$> 10^{10}$	$< -10^{9}$	$> 10^{10}$	.312	5.67	.307	5.67
	β_1	$> 10^{10}$	$> 10^{10}$	$> 10^{9}$	$> 10^{10}$	.782	5.40	.863	5.46
	β_2	$< -10^{10}$	$> 10^{10}$	$< -10^{9}$	$> 10^{10}$	.696	5.52	.696	5.55
	$\lVert\beta\rVert$	$\mathbf{> 10^{10}}$	$\mathbf{> 10^{10}}$	$\mathbf{> 10^{10}}$	$\mathbf{> 10^{10}}$	**1.09**	**9.58**	**1.15**	**9.63**
C9:	β_0	.005	.279	.001	.308	.003	.309	.004	.311
	β_1	$-.002$	.163	$-.005$	.201	$-.004$	.200	$-.005$	.199
	β_2	.001	.165	$-.004$	.204	.003	.198	.002	.198
	$\lVert\beta\rVert$	**.006**	**.363**	**.007**	**.420**	**.006**	**.418**	**.006**	**.419**
C10:	β_0	$-.013$	.284	$-.010$	.315	$-.015$	.314	$-.014$	.314
	β_1	$-.009$	.182	$-.009$	.220	$-.011$	.218	$-.011$	.219
	β_2	.008	.189	.011	.222	.007	.215	.007	.215
	$\lVert\beta\rVert$	**.018**	**.387**	**.018**	**.444**	**.020**	**.439**	**.019**	**.439**
C11:	β_0	.070	1.23	$-.026$	.308	.058	1.26	.053	1.27
	β_1	$-.000$	.268	.005	.214	$-.005$	.351	$-.008$	.354
	β_2	.001	.273	$-.004$	.210	.002	.361	$-.001$	.361
	$\lVert\beta\rVert$	**.070**	**1.29**	**.027**	**.430**	**.059**	**1.36**	**.054**	**1.37**

* $\lVert.\rVert$ stands for the Euclidean norm.

8.1. Drift estimation with heteroskedasticity

In this section, we estimate a constant and a drift on the Standard and Poor's Composite Price Index (SP), 1928-1987. That process is known to involve a large amount of heteroskedasticity and have been used by Gallant, Hsieh and Tauchen (1997) and Dufour and Valéry (2006, 2009) to fit a stochastic volatility model. Here, we are interested in robust **estimation** without modeling the volatility in the disturbance process. The data set consists in a series of 16,127 daily observations of SP_t, then converted in price movements, $y_t = 100[\log(SP_t) - \log(SP_{t-1})]$ and adjusted for systematic calendar effects. We consider a model involving a constant and a drift,

$$y_t = a + bt + u_t, \;\; t = 1, \ldots, 16127, \tag{8.2}$$

and we allow that $\{u_t : t = 1, \ldots, 16127\}$ exhibits stochastic volatility or nonlinear heteroskedasticity of unknown form. White and Breusch-Pagan tests for heteroskedasticity both reject homoskedasticity at 1%.[6]

We compute both the basic *SF* sign-based estimator and the *SHAC* version with the two-step method. They are compared with the LAD and OLS estimates. Then, we redo a similar experiment on two subperiods: on the year 1929 (291 observations) and the last 90 days of 1929, which roughly corresponds to the four last months of 1929 (90 observations). Due to the financial crisis, one may expect data to involve an extreme amount of heteroskedasticity in that period of time. We wonder at which point that heteroskedasticity can bias the subsample estimates. The Wall Street crash occurred between October, 24th (*Black Thursday*) and October, 29th (*Black Tuesday*). Hence, the second subsample corresponds to the period just before the crash (September), the crash period (October) and the early beginning of the Great Depression (November and December). Heteroskedasticity tests reject homoskedasticity for both subsamples.[7]

In Table 3, we report estimates and recall the 95% confidence intervals for a and b obtained by the finite-sample sign-based method (*SF* and *SHAC*);[8] and by moving block bootstrap (LAD and OLS). The entire set of sign-based estimators is reported, *i.e.*, all the minimizers of the sign objective function.

First, OLS estimates are just reported for comparison sake, even if they estimate different quantities as LAD/sign estimators, and are greatly unreliable in the presence of heteroskedasticity. Presenting the entire sets of sign-based estimators enables us to compare them with the LAD estimator. In this example, LAD and sign-based estimators yield very similar estimates. The value of the LAD estimator is indeed just at the limit of the sets of sign-based estimators. This does not mean that the LAD estimator is included in the set of sign-based estimators, but, there is a sign-based estimator giving the same value as the LAD estimate for a certain individual component (the second component may differs). One easy way to check this is to compare the two objective functions evaluated at the two estimates. For example, in the 90 observation sample, the sign objective function evaluated at the basic sign-estimators is 4.75×10^{-3}, and at the LAD estimate 5.10×10^{-2}; the LAD objective

[6] See Coudin and Dufour (2009): White: 499 (p-value=.000) ; BP: 2781 (p-value=.000).

[7] 1929: White: 24.2, p-values: .000 ; BP: 126, p-values: .000; Sept-Oct-Nov-Dec 1929: White: 11.08, p-values: .004; BP: 1.76, p-values: .18.

[8] see Coudin and Dufour (2009)

Table 3. Constant and drift estimates.

Constant parameter (a)	Whole sample (16120obs)	Subsamples 1929(291obs)	1929(90obs)
Set of basic sign-based estimators (SF)	.062 $[-.007, .105]**$	$(.160, .163)*$ $[-.226, .521]$	$(-.091, .142)$ $[-1.453, .491]$
Set of 2-step sign-based estimators (SHAC)	.062 $[-.007, .106]$	$(.160, .163)$ $[-.135, .443]$	$(-.091, .142)$ $[-1.030, .362]$
LAD	.062 $[.008, .116]$	.163 $[-.130, .456]$	$-.091$ $[-1.223, 1.040]$
OLS	$-.005$ $[-.056, .046]$	.224 $[-.140, .588]$	$-.522$ $[-1.730, .685]$
Drift parameter (b)	$\times 10^{-5}$	$\times 10^{-2}$	$\times 10^{-1}$
Set of basic sign-based estimators (SF)	$(-.184, -.178)$ $[-.676, .486]$	$(-.003, .000)$ $[-.330, .342]$	$(-.097, -.044)$ $[-.240, .305]$
Set of 2-step sign-based estimators (SHAC)	$(-.184, -.178)$ $[-.699, .510]$	$(-.003, .000)$ $[-.260, .268]$	$(-.097, -.044)$ $[-.204, .224]$
LAD	$-.184$ $[-.681, .313]$	.000 $[-.236, .236]$	$-.044$ $[-.316, .229]$
OLS	.266 $[-.228, .761]$	$-.183$ $[-.523, .156]$	.010 $[-.250, .270]$

* Interval of admissible estimators (minimizers of the sign objective function).

** 95% confidence intervals.

function evaluated at the LAD estimate is 210.4 and at one of the sign-based estimates 210.5. Both are close but different.

Finally, two-step sign-based estimators and basic sign-based estimators yield the same estimates. Only confidence intervals differ. Both methods are indeed expected to give different results especially in the presence of linear dependence.

8.2. A robust sign-based estimate of convergence across U.S. states

One field suffering from both a small number of observations and possibly very heterogeneous data is cross-sectional regional data sets. Least squares methods may be misleading because a few outlying observations may drastically influence the estimates. Robust methods are greatly needed in such cases. Sign-based estimators are robust (in a statistical sense) and are naturally associated with a finite-sample inference. In the following, we examine sign-based estimates of the rate of β-convergence between output levels across U.S. States between 1880 and 1988 using Barro and Sala-i-Martin (1991) data.

In the neoclassical growth model, Barro and Sala-i-Martin (1991) estimated the rate of β-convergence between levels of per capita output across the U.S. States for different time periods

Table 4. Summary of regression diagnostics.

Period	Heteroskedasticity*		Nonnormality**		Influential observations**		Possible outliers**	
	Basic eq.	*Eq. Reg. Dum.*						
1880-1900	yes	-	yes	-	yes	yes	no	no
1900-1920	yes	yes	yes	yes	yes	yes	yes (MT)	yes
1920-1930	-	-	-	-	yes	-	no	no
1930-1940	-	-	yes	-	yes	yes	no	no
1940-1950	-	-	-	-	yes	yes	yes (VT)	yes (VT)
1950-1960	-	-	-	yes	yes	yes	yes (MT)	yes (MT)
1960-1970	-	-	-	-	-	-	no	no
1970-1980	-	-	yes	yes	yes	yes	yes (WY)	yes (WY)
1980-1988	yes	-	-	yes	yes	yes	yes (WY)	yes (WY)

* White and Breusch-Pagan tests for heteroskedasticity are performed. If at least one test rejects at 5% homoskedasticity, a "yes" is reported in the table, else a "-" is reported, when tests are both nonconclusive.
** Scatter plots, kernel density, leverage analysis, Studentized or standardized residuals larger than 3, DFbeta and Cooks distance have been performed and lead to suspicions for nonnormality, outlier or high influential observation presence.

between 1880 and 1988. They used nonlinear least squares to estimate equations of the form

$$(1/T)\ln(y_{i,t}/y_{i,t-T}) = a - [\ln(y_{i,t-T})] \times [(1-e^{-\beta T})/T] + x_i'\delta + \varepsilon_i^{t,T},$$

$i = 1,\ldots,48,\ T = 8,\ 10$ or $20,\ \ t = 1900,\ 1920,\ 1930,\ 1940, 1950,\ 1960,\ 1970,\ 1980,\ 1988.$ Their *basic equation* does not include any other variables but they also consider a specification with regional dummies (*Eq. with reg. dum.*). The *basic equation* assumes that the 48 States share a common per capita level of personal income at steady state while the second specification allows for regional differences in steady state levels. Their regressions involve 48 observations and are run for each 20-year or 10-year period between 1880 and 1988. Their results suggest a β-convergence at a rate somewhat above 2% a year but their estimates are not stable across subperiods, and vary greatly from -.0149 to .0431 (for the *basic equation*). This instability is expected because of the succession of troubles and growth periods in the last century. However, they may also be due to particular observations behaving like outliers and influencing the least-squares estimates. A survey of potential data problem is performed and regression diagnostics are summarized in Table 4. It suggests the presence of highly influential observations in all the periods but one. Outliers are clearly identified in periods 1900-1920, 1940-1950, 1950-1960, 1970-1980 and 1980-1988.

These two effects are probably combined. We wonder which part of that variability is really due to business cycles and which part is only due to the non-robustness of least-squares methods. Further, we would like to have a stable estimate of the rate of convergence at steady state. For this, we use robust sign-based estimation with $D_S\big(\beta,(X'X)^{-1}\big)$. We consider the following linear

Table 5. Regressions for personal income across U.S. States, 1880-1988. estimates of β

Period	*Basic equation*		*Equation with regional dummies*	
	SIGN	NLLS***	SIGN	NLLS***
1880 – 1900	.0012	.0101	.0016	.0224
	$[-.0068, .0123]^*$	$[.0058, .0532]^{**}$	$[-.0123, .0211]$	$[.0146, .0302]$
1900 – 1920	.0184	.0218	.0163	.0209
	$[.0092, .0313]$	$[.0155, .0281]$	$[-.0088, .1063]$	$[.0086, .0332]$
1920 – 1930	$-.0147$	$-.0149$	$-.0002$	$-.0122$
	$[-.0301, .0018]$	$[-.0249, -.0049]$	$[-.0463, .0389]$	$[-.0267, .0023]$
1930 – 1940	.0130	.0141	.0152	.0127
	$[.0043, .0234]$	$[.0082, .0200]$	$[-.0189, .0582]$	$[.0027, .0227]$
1940 – 1950	.0364	.0431	.0174	.0373
	$[.0291, .0602]$	$[.0372, .0490]$	$[.0083, .0620]$	$[.0314, .0432]$
1950 – 1960	.0195	.0190	.0140	.0202
	$[.0084, .0352]$	$[.0121, .0259]$	$[-.0044, .0510]$	$[.0100, .0304]$
1960 – 1970	.0289	.0246	.0230	.0131
	$[.0099, .0377]$	$[.0170, .0322]$	$[-.0112, .0431]$	$[.0047, .0215]$
1970 – 1980	.0181	.0198	.0172	.0119
	$[.0021, .0346]$	$[-.0315, .0195]$	$[-.0131, .0739]$	$[-.0273, .0173]$
1980 – 1988	$-.0081$	$-.0060$	$-.0059$	$-.0050$
	$[-.0552, .0503]$	(.0130)	$[-.0472, .1344]$	(.0114)

* Projection-based 95% CI.

** Asymptotic 95% CI.

*** Estimates from Barro and Sala-i-Martin (1991).

equation:

$$(1/T)\ln(y_{i,t}/y_{i,t-T}) = a + \gamma[\ln(y_{i,t-T})] + x_i'\delta + \varepsilon_i^{t,T} \tag{8.3}$$

where x_i's contain regional dummies when included, and we compute Hodges-Lehmann estimate for $\beta = -(1/T)\ln(\gamma T + 1)$ for both specifications. We also provide 95%-level projection-based CI, asymptotic CI and projection-based p-value functions for the parameter of interest. Results are presented in Table 5 where Barro and Sala-i-Martin (1991) NLLS results are reported.

Sign estimates are more stable than least-squares ones. They vary between $[-.0147, .0364]$ whereas least-squares estimates vary between $[-.0149, .0431]$. This suggests that at least 12% of the least-squares estimates variability between sub-periods are only due to the non-robustness of least-squares methods. In all cases but two, sign-based estimates are lower (in absolute values) than the NLLS ones. Consequently, we incline to a lower value of the stable rate of convergence.

In graphics 6(a)-8(f) [see Appendix B], projection-based p-value functions and optimal concentrated sign-statistics are presented for each *basic equation* over the period 1880-1988. The optimal concentrated sign-based statistic reports the minimal value of D_S for a given β (letting a varying). The projection-based p-value function is the maximal simulated p-value for a given β over admissible values of a. Those functions enable us to perform tests on β. 95% projection based confidence intervals for β presented in Table 5 are obtained by cutting the p-value function with the $p = .05$

line. The sign estimate reaches the highest p-value. Remark that contrary to asymptotic methods, the estimator is not at the middle point of any confidence interval. Besides, the p-value function gives some hint on the degree of precision. The β parameter seems precisely estimated in the period 30-40 [see graphic 7(b)], whereas in the period 80-88, the same parameter is less precisely estimated and the p-value function leads to a wider confidence intervals [see graphic 8(f)].

9. Conclusion

In this paper, we have introduced inference tools that can be associated with the Monte Carlo based system presented in Coudin and Dufour (2009): the p-value function (and its individual projected versions) which gives a visual summary of all the inference available on a particular parameter, and Hodges-Lehmann-type sign-based estimators. The p-value function associates to each value of the parameter vector a confidence degree. It extends the confidence distribution concept to multidimensional parameters and relies on a reinterpretation of the Fisher fiducial distributions. Parameter values least rejected by tests (given the sample realization and the sample size) constitute Hodges-Lehmann sign-based estimators. Those estimators are associated with the highest p-value. Hence, they are derived without referring to asymptotic conditions through the analogy principle. However, they turn out to be equivalent (in probability) to usual GMM estimators based on signs.

We then derived some general properties of sign-based estimators (invariance, median unbiasedness) and conditions under which consistency and asymptotic normality hold. In particular, we showed that sign-based estimators do require less assumptions on moment existence of the disturbances than usual LAD asymptotic theory. Simulation studies indicate that the proposed estimators are accurate in classical setups and more reliable than usual methods (LS, LAD) when heterogeneity or nonlinear dependence is present in the error term even in cases that may cause LAD or OLS consistency failure. Despite the programming complexity of sign-based methods, we recommend combining sign-based estimators to the Monte Carlo sign-based method of inference when an amount of heteroskedasticity is suspected in the data and when the number of available observations is small. Finally, we presented two illustrative applications of such cases. In the first one, we estimated a drift parameter on the Standard and Poor's Composite Price Index, using the 1928-1987 period as well as shorter subperiods. In the second one, we provided robust estimates for the β-convergence between the levels of per capita personal income across U.S. States between 1880 and 1988.

Appendix

A. Proofs

Proof of Proposition 4.1. Let D_S be a sign-based statistic of the form presented in equation (4.4). The term Ω_n is omitted for simplicity. We show that the sets M_1 and M_2 are equal with probability one. First, we show that if $\hat{\beta} \in M_2$ then it belongs to M_1. Second, we show that if $\hat{\beta}$ does not belong to M_2, neither it belongs to M_1.
If $\hat{\beta} \in M_2$ then,

$$D_S(\hat{\beta}) \leq D_S(\beta),\ \forall \beta \in \Theta, \tag{A.1}$$

hence

$$\mathsf{P}_\beta[D_S(\hat{\beta}) \leq D_S(\beta)] = 1, \quad \forall \beta \in \Theta \tag{A.2}$$

and $\hat{\beta}$ maximizes the p-value. Conversely, if $\hat{\beta}$ does not belong to M_1, there is a non negligible Borel set, say A, such that $D_S(\beta) < D_S(\hat{\beta})$ on A for some β. Then, as $\bar{F}(x)$, the distribution function of D_S is an increasing function and A is non negligible, and since $\bar{F}$ is independent of β (Assumption 4.1),

$$\bar{F}\big(D_S(\beta)\big) < \bar{F}\big(D_S(\hat{\beta})\big). \tag{A.3}$$

Finally, equation A.3 can be written in terms of p-values

$$\mathrm{p}(\beta) > \mathrm{p}(\hat{\beta}), \tag{A.4}$$

which implies that $\hat{\beta}$ does not belong to M_2. □

Proof of Proposition 5.2. Consider $\hat{\beta}(y,X,u)$ the solution of problem (4.4) which is assumed to be unique, let $\bar{\beta}$ be the true value of the parameter β and suppose that $u \sim -u$. Equation (5.4) implies that

$$\hat{\beta}(u,X,u) = -\hat{\beta}(-u,X,u)$$

where both problems are assumed to have a single solution. Hence, conditional on X, we have

$$u \sim -u \Rightarrow \hat{\beta}(u,X,u) \sim -\hat{\beta}(-u,X,u) \Rightarrow \mathrm{Med}\big(\hat{\beta}(u,X,u)\big) = 0. \tag{A.5}$$

Moreover, equation (5.5) implies that

$$\begin{aligned} \hat{\beta}(y,X,u) &= \hat{\beta}(y - X\bar{\beta}, X, u) + \bar{\beta} \\ &= \hat{\beta}(u,X,u) + \bar{\beta}. \end{aligned} \tag{A.6}$$

Finally, (A.5) and (A.6) entail $\mathrm{Med}(\hat{\beta}(y,X,u) - \bar{\beta}) = 0$. □

Proof of Theorem 6.1. We consider the stochastic process $W = \{W_t = (y_t,\ x_t') : \Omega \to \mathbb{R}^{p+1}\}_{t=1,2,\ldots}$

defined on the probability space $(\Omega, \mathscr{F}, \mathsf{P})$. We denote

$$\begin{aligned} q_t(W_t,\beta) &= \left[q_{t1}(W_t,\beta),\ldots,q_{tp}(W_t,\beta)\right]' \\ &= \left[s(y_t - x_t'\beta)x_{t1},\ldots,s(y_t - x_t'\beta)x_{tp}\right]',\ t = 1,\ldots,n. \end{aligned}$$

The proof of consistency follows four classical steps. First, $\frac{1}{n}\sum_t q_t(W_t,\beta) - \mathsf{E}[q_t(W_t,\beta)]$ is shown to converge in probability to zero for all $\beta \in \Theta$ (**pointwise convergence**). Second, that convergence is extended to a **weak uniform convergence**. Third, we adapt to our setup the **consistency theorem** of extremum estimators of Newey and McFadden (1994). Fourth, consistency is entailed by the **optimum uniqueness** that results from the identification conditions.

Pointwise convergence. The mixing property 6.1 on W is exported to $\{q_{tk}(W_t,\beta),\ k = 1,\ldots,p\}_{t=1,2,\ldots}$. Hence, $\forall \beta \in \Theta$, $\forall k = 1,\ldots,p$, $\{q_{tk}(W_t,\beta)\}$ is an $\alpha-$mixing process of size $r/(1-r)$. Moreover, condition 6.2 entails $\mathsf{E}|q_{tk}(W_t,\beta)|^{r+\delta} < \infty$ for some $\delta > 0$, for all $t \in \mathbb{N}$, $k = 1,\ldots,p$. Hence, we can apply Corollary 3.48 of White (2001) to $\{q_{tk}(W_t,\beta)\}_{t=1,2,\ldots}$. It follows $\forall \beta \in \Theta$,

$$\frac{1}{n}\sum_{t=1}^{n} q_{tk}(W_t,\beta) - \mathsf{E}[q_{tk}(W_t,\beta)] \xrightarrow{p} 0 \ \ k = 1,\ldots,p,$$

Uniform Convergence. We check conditions A1, A6, B1, B2 of Andrews (1987)'s generic weak law of large numbers (GWLLN). A1 and B1 are our conditions 6.3 and 6.1. Then, Andrews defines

$$\begin{aligned} q_{ik}^{H}(W_i,\beta,\rho) &= \sup_{\hat\beta \in B(\beta,\rho)} q_{ik}(W_i,\hat\beta), \\ q_{Lik}(W_i,\beta,\rho) &= \inf_{\hat\beta \in B(\beta,\rho)} q_{ik}(W_i,\hat\beta), \end{aligned}$$

where $B(\beta,\rho)$ is the open ball around β of radius ρ. His condition B2 requires that $q_{tk}^{H}(W_t,\beta,\rho)$, $q_{Ltk}(W_t,\beta,\rho)$ and $q_{tk}(W_t)$ are random variables; $q_{tk}^{H}(.,\beta,\rho)$, $q_{Ltk}(.,\beta,\rho)$ are measurable functions from $(\Omega,\mathscr{P},\mathscr{F})$ to $(\mathbb{R},\mathscr{B})$, $\forall t$, $\beta \in \Theta$, ρ, where $\mathscr{B}$ is the Borel σ-algebra on $\mathbb{R}$ and finally, that $\sup_t \mathsf{E}q_{tk}(W_t)^{\xi} < \infty$ with $\xi > r$. Those points are derived from the mixing condition 6.1 and condition 6.2 which insures measurability and provides bounded arguments.

The last condition (A6) to check requires the following: Let μ be a σ-finite measure that dominates each one of the marginal distributions of W_t, $t = 1,2\ldots$. Let $p_t(w)$ be the density of W_t w.r.t. μ, $q_{tk}(W_t,\ \beta)p_t(W_t)$ is continuous in β at $\beta = \beta^*$ uniformly in t a.e. w.r.t. μ, for each $\beta^* \in \Theta$, $q_{tk}(W_t,\ \beta)$ is measurable w.r.t. the Borel measure for each t and each $\beta \in \Theta$, and $\int \sup_{t\geq 0,\ \beta\in\Theta} |q_{tk}(W,\ \beta)|p_t(w)d\mu(w) < \infty$. As u_t is continuously distributed uniformly in t [Assumption 6.4 (2)], we have $\mathsf{P}_t[u_t = x_t\beta] = 0$, $\forall \beta$, uniformly in t. Then, q_{tk} is continuous in β everywhere except on a P_t-negligible set. Finally, since q_{tk} is L_1-bounded and uniformly integrable, condition A6 holds.

The generic law of large numbers (GWLLN) implies:

(a) $\qquad \dfrac{1}{n}\displaystyle\sum_{i=0}^{n} \mathsf{E}[q_t(W_t,\ \beta)]$ is continuous on Θ uniformly over $n \geq 1$,

(b) $$\sup_{\beta\in\Theta}\left|\frac{1}{n}\sum_{t=0}^{n} q_t(W_t,\ \beta) - \mathsf{E}q_t(W_t,\ \beta)\right| \to 0$$
as $n \to \infty$ in probability under P.

The **Consistency Theorem** consists in an extension of Theorem 2.1 of Newey and McFadden (1994) on extremum estimators. The steps of the proof are the same but the limit problem slightly differs. For simplicity, the true value is taken to be 0. First, the generic law of large numbers entails that

$$\lim_{n\to\infty}\frac{1}{n}\sum_t \mathsf{E}[s(u_t - x_t'\beta)x_{tk}] \text{ is continuous on } \Theta, k = 1,\ldots,p. \tag{A.7}$$

Let us define

$$Q_n^k(\beta) = \frac{1}{n}\left|\sum_{t=1}^{n} x_{kt}s(u_t - x_t'\beta)\right|,\ k = 1,\ldots,p,$$

$$Q_n^{Ek}(\beta) = \frac{1}{n}\left|\sum_{t=1}^{n} \mathsf{E}[x_{kt}s(u_t - x_t'\beta)]\right|,\ k = 1,\ldots,p.$$

We consider $\{\beta_n\}_{n\geq 1}$ a sequence of minimizers of the objective function of the non-weighted sign-based estimator

$$\frac{1}{n^2}\sum_{k=1}^{p}\left(\sum_t x_{kt}s(u_t - x_t'\beta)\right)^2 = \sum_k [Q_n^k(\beta)]^2.$$

Then for all $\varepsilon > 0$, $\delta > 0$ and $n \geq N_0$, we have:

$$\mathsf{P}\left[\sum_k [Q_n^k(\beta_n)]^2 < \sum_k [Q_n^k(0)]^2 + \varepsilon/3\right] \geq 1 - \delta. \tag{A.8}$$

Uniform weak convergence of Q_n^k to Q_n^{Ek} at β_n implies:

$$[Q_n^{Ek}(\beta_n)]^2 < [Q_n^k(\beta_n)]^2 + \varepsilon/3p,\ k = 1,\ldots,p,\ \text{with probability approaching one as } n \to \infty, \tag{A.9}$$

hence,

$$\sum_k [Q_n^{Ek}(\beta_n)]^2 < \sum_k [Q_n^k(\beta_n)]^2 + \varepsilon/3,\ \text{with probability approaching one as } n \to \infty. \tag{A.10}$$

With the same argument, at $\beta = 0$

$$\sum_k [Q_n^k(0)]^2 < \sum_k [Q_n^{Ek}(0)]^2 + \varepsilon/3,\ \text{with probability approaching one as } n \to \infty. \tag{A.11}$$

Using (A.10), (A.8) and (A.11) in turn, this entails

$$\sum_k [Q_n^{Ek}(\beta_n)]^2 < \sum_k [Q_n^{Ek}(0)]^2 + \varepsilon,\ \text{with probability approaching one as } n \to \infty. \tag{A.12}$$

This holds for any ε, with probability approaching one. Let $\mathbf{N}$ be any open subset of Θ containing 0. As $\Theta \cap \mathbf{N}^c$ is compact and $\lim_n \sum_k [Q_n^{*k}(\beta)]^2$ is continuous (A.7),

$$\exists \beta^* \in \Theta \cap \mathbf{N}^c \text{ such that } \sup_{\beta \in \Theta \cap \mathbf{N}^c} \lim_n \sum_k [Q_n^{Ek}(\beta)]^2 = \lim_n \sum_k [Q_n^{Ek}(\beta^*)]^2.$$

Provided that 0 is the unique minimizer, we have:

$$\lim_n \sum_k [Q_n^{Ek}(\beta^*)]^2 > \lim_n \sum_k [Q_n^{Ek}(0)]^2, \text{ with probability one }.$$

Hence, setting

$$\varepsilon = \frac{1}{2}\left\{\lim_n \sum_k [Q_n^{Ek}(\beta^*)]^2\right\},$$

it follows that, with probability close to one,

$$\lim_n \sum_k [Q_n^{Ek}(\beta_n)]^2 < \frac{1}{2}\left[\lim_n \sum_k [Q_n^{Ek}(\beta^*)]^2 + \lim_n \sum_k [Q_n^{Ek}(0)]^2\right] < \sup_{\beta \in \Theta \cap \mathbf{N}^c} \lim_n \sum_k [Q_n^{Ek}(\beta)]^2.$$

Hence, $\beta_n \in \mathbf{N}$. As this holds for any open subset $\mathbf{N}$ of Θ we conclude on the convergence of β_n to 0.

For **identification**, the uniqueness of the minimizer of the sign-objective function is insured by the set of identification conditions 2.2, 6.5, 6.4, 6.6. These conditions and consequently the proof, are close to those of Weiss (1991) and Fitzenberger (1997) for the LAD and quantile estimators. We wish to show that the limit problem does not admit another solution. When $\bar{\Omega}_n(\beta)$ defines a norm for each β (condition 6.6), this assertion is equivalent to

$$\lim_{n\to\infty} \mathsf{E}\left[\frac{1}{n}\sum_t s(u_t - x_t'\delta)x_i\right] = 0 \Rightarrow \delta = 0,\ \delta \in \mathbb{R}^p, \tag{A.13}$$

and

$$\lim_{n\to\infty} \left|\mathsf{E}\left[\frac{1}{n}\sum_t s(u_t - x_t'\delta)x_t'\delta\right]\right| = 0 \Rightarrow \delta = 0,\ \delta \in \mathbb{R}^p. \tag{A.14}$$

Let $A(\delta) = \mathsf{E}[\frac{1}{n}\sum_t s(u_t - x_t'\delta)x_t | x_1, \ldots, x_n]$. Then,

$$\mathsf{E}[A(\delta)] = \mathsf{E}\left[\frac{1}{n}\sum_t s(u_t - x_t'\delta)x_t\right] = \mathsf{E}\left\{\mathsf{E}\left[\frac{1}{n}\sum_t s(u_t - x_t'\delta)x_t | x_1, \ldots, x_n\right]\right\}.$$

Note that

$$\mathsf{E}[s(u_t - x_t'\delta)|x_1, \ldots, x_n] = 2\left[\frac{1}{2} - \int_{-\infty}^{x_t'\delta} f_t(u|x_1, \ldots, x_n)du\right] = -2\int_0^{x_t'\delta} f_t(u|x_1, \ldots, x_n)du]$$

Hence $A(\delta)$ can be developed for $\tau > 0$ as

$$
\begin{aligned}
A(\delta) \quad = \quad & \frac{2}{n}\sum x_t'\delta\left\{ I_{\{|x_t'\delta|>\tau\}}\left[I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta} -f_t(u|x_1,\ldots,x_n)du \right.\right. \\
& \left. +I_{\{x_t'\delta\le 0\}}\int_{x_t'\delta}^{0} f_t(u|x_1,\ldots,x_n)du \right] \\
& +I_{\{|x_t'\delta|\le\tau\}}\left[I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta} -f_t(u|x_1,\ldots,x_n)du \right. \\
& \left.\left. +I_{\{x_t'\delta\le 0\}}\int_{x_t'\delta}^{0} f_t(u|x_1,\ldots,x_n)du \right]\right\}.
\end{aligned}
$$

Then,

$$
\begin{aligned}
& \mathsf{E}[A(\delta)] = \mathsf{E}\left\{\frac{2}{n}\sum x_t'\delta\left[I_{\{|x_t'\delta|>\tau\}}\left(I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta} -f_t(u|x_1,\ldots,x_n)du \right.\right.\right. \\
& \left. +I_{\{x_t'\delta\le 0\}}\int_{x_t'\delta}^{0} f_t(u|x_1,\ldots,x_n)du \right) \\
& +I_{\{|x_t'\delta|\le\tau\}}(I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta} -f_t(u|x_1,\ldots,x_n)du \\
& \left.\left. +\ I_{\{x_t'\delta\le 0\}}\int_{x_t'\delta}^{0} f_t(u|x_1,\ldots,x_n)du)\right]\right\}.
\end{aligned}
$$

Remark that each term in this sum is negative. Hence, $s(\mathsf{E}[A(\delta)]) \le 0$ and $|\mathsf{E}[A(\delta)]| = -\mathsf{E}[A(\delta)]$, and

$$
\begin{aligned}
|\mathsf{E}(A)| \quad = \quad & \mathsf{E}\left[\frac{2}{n}\sum x_t'\delta I_{\{|x_t'\delta|>\tau\}}\left(I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta} f_t(u|x_1,\ldots,x_n)du \right.\right. \\
& \left.\left. -I_{\{x_t'\delta\le 0\}}\int_{x_t'\delta}^{0} f_t(u|x_1,\ldots,x_n)du \right)\right] \\
& +\mathsf{E}\left[\frac{2}{n}\sum x_t'\delta I_{\{|x_t'\delta|\le\tau\}}\left(I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta} f_t(u|x_1,\ldots,x_n)du \right.\right. \\
& \left.\left. -I_{\{x_t'\delta\le 0\}}\int_{x_t'\delta}^{0} f_t(u|x_1,\ldots,x_n)du \right)\right] \\
\ge \quad & \mathsf{E}\left[\frac{2}{n}\sum I_{\{|x_t'\delta|>\tau\}}\left(x_t'\delta I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta} f_t(u|x_1,\ldots,x_n)du \right.\right. \\
& \left.\left. -x_t'\delta I_{\{x_t'\delta\le 0\}}\int_{x_t'\delta}^{0} f_t(u|x_1,\ldots,x_n)du \right)\right] \qquad \text{(A.15)} \\
\ge \quad & \mathsf{E}\left\{\frac{2}{n}\sum I_{\{|x_t'\delta|>\tau\}}\left[x_t'\delta I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta} f_t(u|x_1,\ldots,x_n)du \right.\right.
\end{aligned}
$$

$$- \quad x_t'\delta I_{\{x_t'\delta \leq 0\}} \int_{x_t'\delta}^{0} f_t(u|x_1,\ldots,x_n)du \Big] [f_t(0|x_1,\ldots,x_n) > f_L] p_1 \Big\} \tag{A.16}$$

$$\geq \quad p_1 \mathsf{E}\left\{ \frac{2}{n} \sum I_{\{|x_t'\delta|>\tau\}} \tau f_L d | f_t(0|x_1,\ldots,x_n) > f_L \right\}, \tag{A.17}$$

$$\geq \quad \tau p_1 f_L d \frac{2}{n} \sum P[|x_t'\delta| > \tau | f_t(0|x_1,\ldots,x_n) > f_L)]. \tag{A.18}$$

To obtain inequation (A.15), just remark that each term is positive. For the inequation (A.16) we use condition 6.4. For inequation (A.17) we minorate $|x_i'\delta|$ by τ and each integrals by $f_L d_1$ where $d_1 = \min(\tau, d/2)$. Condition 6.5 enables us to conclude, by taking the limit,

$$\lim_{n\to\infty} |\mathsf{E}[A(\delta)]| \geq 2\tau p_1 f_L d \times \liminf_{n\to\infty} \mathsf{P}[|x_i'\delta| > \tau | f_i(0|x_1,\ldots,x_n) > f_L] > 0, \quad \forall \delta > 0,$$

hence, we conclude on the uniqueness of the minimum, which was the last step to insure consistency of the sign-based estimators. □

Proof of Theorem 6.2. We prove Theorem 6.2 on asymptotic normality. We consider the sign-based estimator $\hat{\beta}(\Omega_n)$ where Ω_n stands for any $p \times p$ positive definite matrix. We apply Theorem 7.2 of Newey and McFadden (1994), which allows to deal with noncontinuous and nondifferentiable objective functions for finite n. Thus, we stand out from usual proofs of asymptotic normality for the LAD or the quantile estimators, for which the objective function is at least continuous. In our case, only the limit objective function is continuous (see the consistency proof). The proof is separated in two parts. First, we show that $L(\beta)$ as defined in equation (6.3) is the derivative of $\lim_{n\to\infty} \frac{1}{n} \sum_t \mathsf{E}\big[s\big(u_t - x_t'(\beta - \bar{\beta})\big)x_t\big]$. Then, we check the conditions for applying Theorem 7.2 of Newey-McFadden.

The consistency proof (generic law of large numbers) implies that

$$\frac{1}{n} \sum_{t=0}^{n} \mathsf{E}\big[s\big(u_t - x_t'(\beta - \bar{\beta})\big)x_t\big] \tag{A.19}$$

is continuous on Θ uniformly over n. Moreover condition 6.2 specifies that X is $L^{2+\delta}$ bounded. As the $f_t(\lambda|x_1,\ldots,x_n)$ are bounded by f_U uniformly over n and λ (condition 6.8), dominated convergence allows us to write that

$$\frac{\partial}{\partial \beta} \mathsf{E}\big[x_t s\big(u_t - x_t'(\beta - \bar{\beta})\big)\big] = \mathsf{E}\big[x_t x_t' f_t\big(x_t'(\beta - \bar{\beta})|x_1,\ldots,x_n\big)\big]. \tag{A.20}$$

And, these conditions imply that

$$L_n(\beta) = \frac{1}{n} \sum_{t=1}^{n} \mathsf{E}\big[x_t x_t' f_t\big(x_t'(\beta - \bar{\beta})|x_1,\ldots,x_n\big)\big] \tag{A.21}$$

converges uniformly in β to $L(\beta)$. Uniform convergence entails that $\lim_n \frac{1}{n}\sum_{t=0}^{n} \mathsf{E}\big[s\big(u_t - x_t'(\beta - \bar{\beta})\big)x_t\big]$ is differentiable with derivative $L(\beta)$.
We now apply Theorem 7.2 of Newey and McFadden (1994) which presents asymptotic normality of a minimum distance consistent estimator with nonsmooth objective function and weight matrix $\Omega_n \xrightarrow{p} \Omega$ symmetric positive definite. Thus, under conditions for consistency (2.2, 6.1-6.6), we have to check that the following conditions hold:

(i) zero is attained at the limit by $\bar{\beta}$;

(ii) the limiting objective function is differentiable at $\bar{\beta}$ with derivative $L(\bar{\beta})$ such that $L(\bar{\beta})\Omega L(\bar{\beta})'$ is nonsingular;

(iii) $\bar{\beta}$ is an interior point of Θ;

(iv) $\sqrt{n}Q_n(\bar{\beta}) \to \mathcal{N}(0,J)$;

(v) for any $\delta_n \to 0$, $\sup_{||\beta-\bar{\beta}||} \sqrt{n}||Q_n(\beta) - Q_n(\bar{\beta}) - \mathsf{E}Q(\beta)||/(1+\sqrt{n}||\beta-\bar{\beta}||) \xrightarrow{p} 0$.

Condition (i) is fulfilled by the moment condition 2.2. Condition (ii) is fulfilled by the first part of our proof and condition 6.10. Then, Condition (iii) is implied by 6.3. Using the mixing specification 6.9 of $\{u_t, X_t\}_{t=1,2,...}$ and conditions 2.2, 6.2, 6.7 and 6.11, 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):

$$\sqrt{n}J_n^{-1/2}Q_n(\bar{\beta}) \to N(0,I_p) \tag{A.22}$$

where $J_n = \text{var}\left[\frac{1}{\sqrt{n}}\sum_1^n s(u_i)x_i\right]$. Finally, condition (v) can be viewed as a stochastic equicontinuity condition and is easily derived from the uniform convergence [see McFadden remarks on condition (v)]. Hence, $\hat{\beta}(\Omega_n)$ is asymptotically normal

$$\sqrt{n}S_n^{-1/2}\big[\hat{\beta}(\Omega_n) - \bar{\beta}\big] \to \mathcal{N}(0,I_p).$$

The asymptotic covariance matrix S is given by the limit of

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

When choosing $\Omega_n = \hat{J}_n^{-1}$ a consistent estimator of J_n^{-1}, S_n can be simplified:

$$\sqrt{n}S_n^{-1/2}\big[\hat{\beta}(\hat{J}_n^{-1}) - \bar{\beta}\big] \to \mathcal{N}(0,I_p)$$

with

$$S_n = [L_n(\bar{\beta})\hat{J}_n^{-1}L_n(\bar{\beta})]^{-1}.$$

When the mediangale Assumption (2.1) holds, we find usual results on sign-based estimators. $\hat{\beta}(I_p)$

and $\hat{\beta}((X'X)^{-1})$ are asymptotically normal with asymptotic covariance matrix

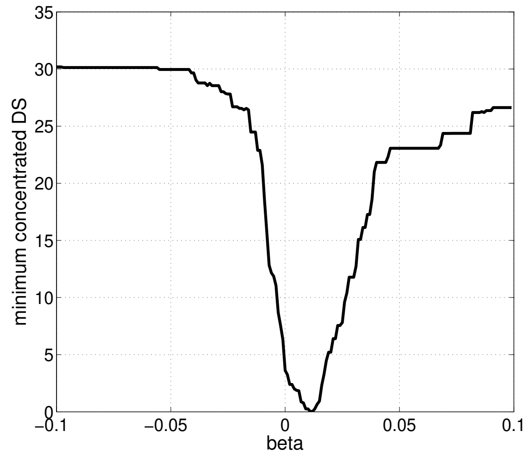
$$\lim_{n\to\infty} S_n = \lim_{n\to\infty} \frac{n^2}{4} \left[\sum_t \mathsf{E}\left[x_t x_t' f_t(0|X)\right]\right]^{-1} \mathsf{E}(x_t x_t') \left[\sum_i \mathsf{E}\left[x_t x_t' f_t(0|X)\right]\right]^{-1}.$$

□

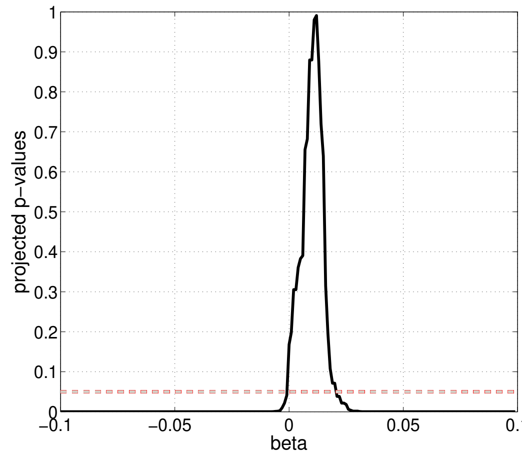
B. Convergence data: concentrated statistics and p-values

This appendix contains graphics of concentrated sign-based statistics and projected p-values for the β parameter in the Barro and Sala-i-Martin application.

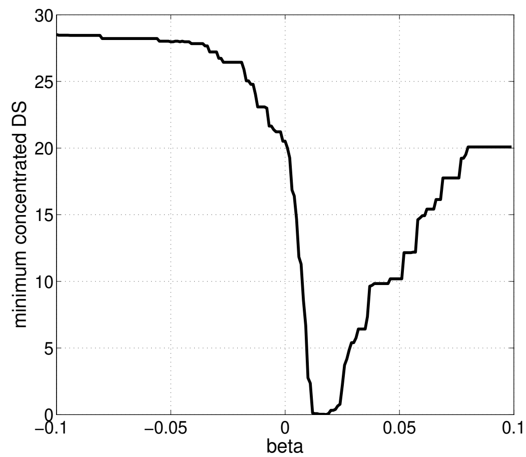
Figure 5. Concentrated statistics and projected *p*-values (1880-1930)



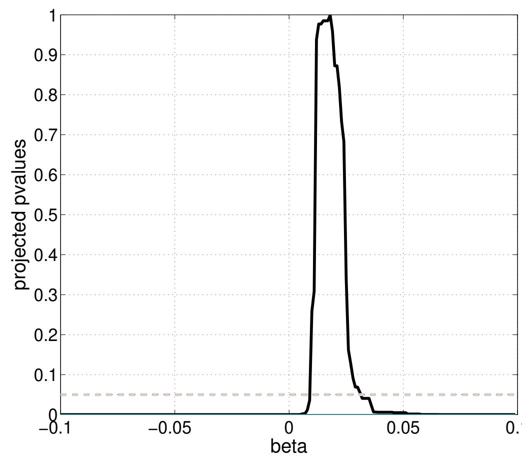
(a) Basic equation: 1880-1900: concentrated DS



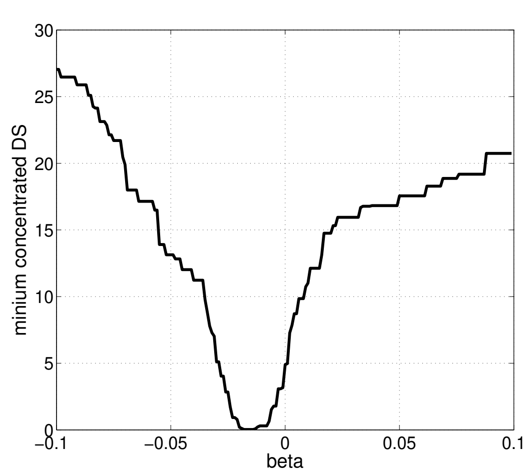
(b) Basic equation: 1880-1900: projected *p*-value



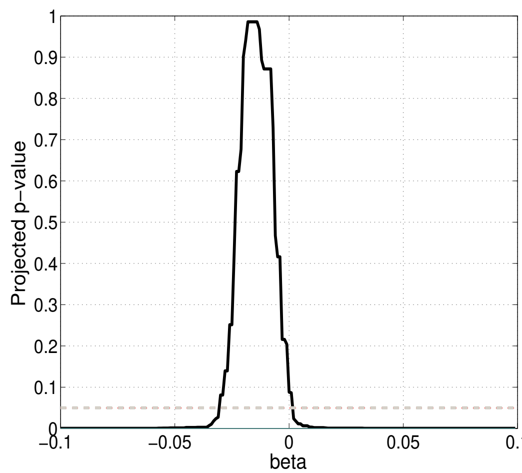
(c) Basic equation: 1900-20: concentrated DS



(d) Basic equation: 1900-20: projected *p*-value

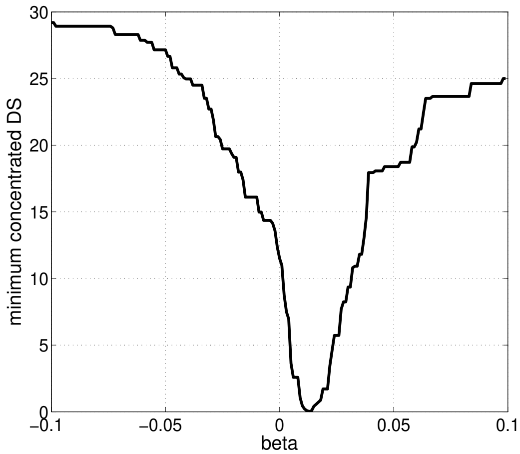


(e) Basic equation: 1920-30: concentrated DS

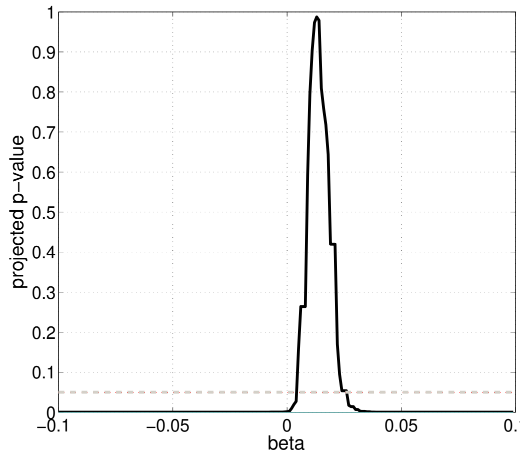


(f) Basic equation: 1920-30: projected *p*-value

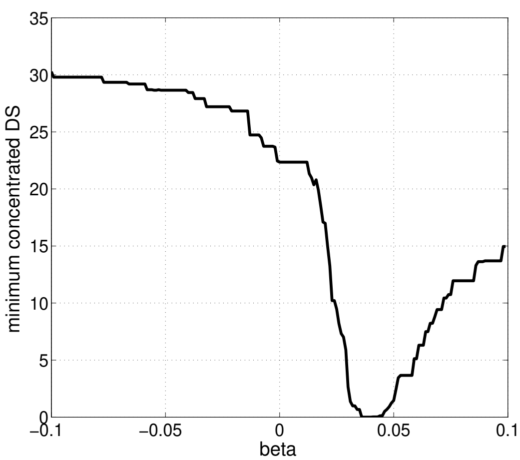
Figure 6. Concentrated statistics and projected p-values (1930-1960)



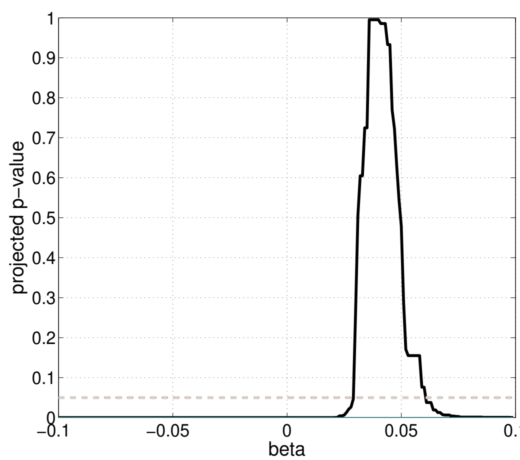
(a) Basic equation: 1930-40: concentrated DS



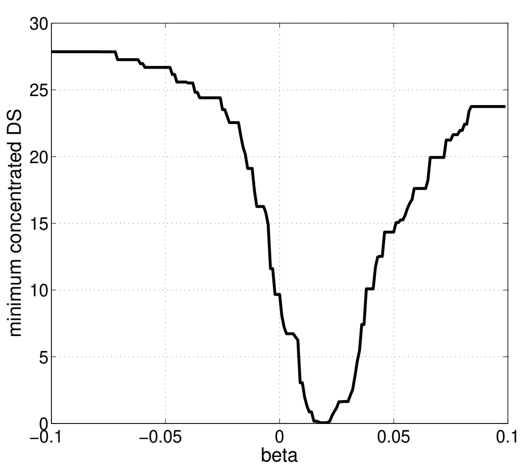
(b) Basic equation: 1930-40: projected p-value



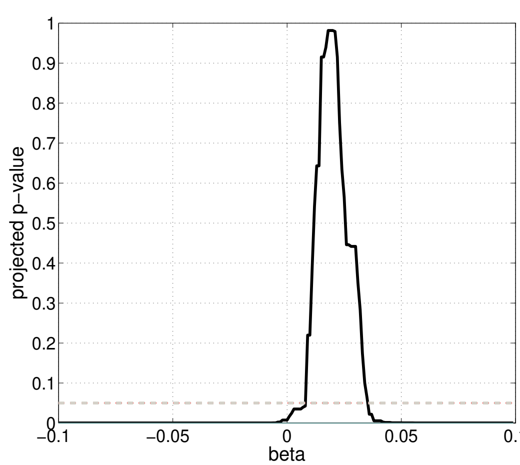
(c) Basic equation: 1940-50: concentrated DS



(d) Basic equation: 1940-50: projected p-value

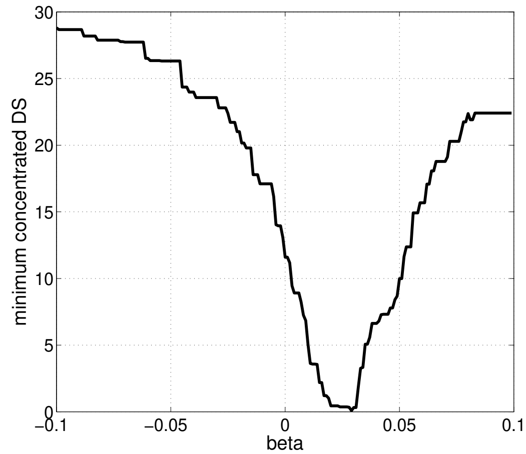


(e) Basic equation: 1950-60: concentrated DS

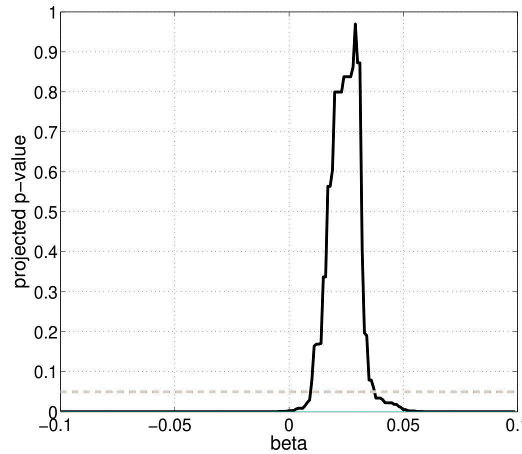


(f) Basic equation: 1950-60: projected p-value

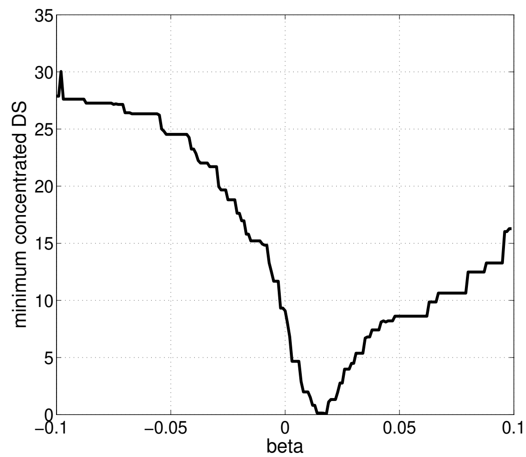
Figure 7. Concentrated statistics and projected p-values (1960-1988)



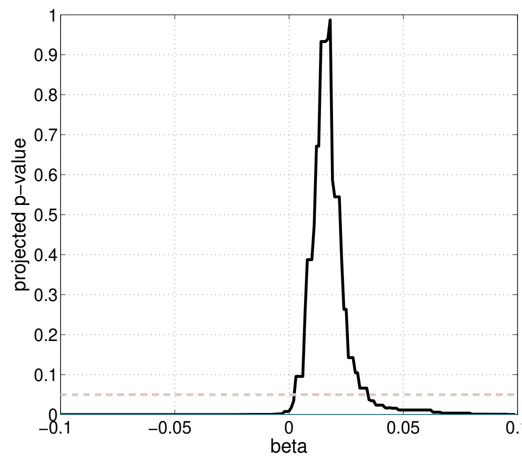
(a) Basic equation: 1960-70: concentrated DS



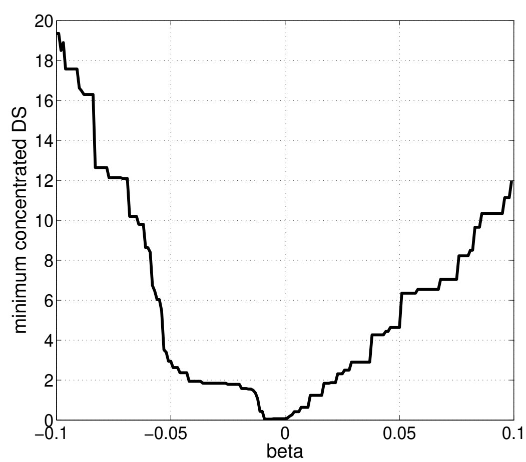
(b) Basic equation: 1960-70: projected p-value



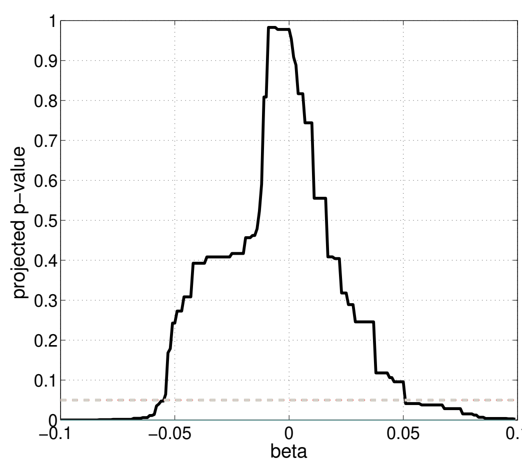
(c) Basic equation: 1970-80: concentrated DS



(d) Basic equation: 1970-80: projected p-value



(e) Basic equation: 1980-88: concentrated DS



(f) Basic equation: 1980-88: projected p-value

References

Abdelkhalek, T. and Dufour, J.-M. (1998), 'Statistical inference for computable general equilibrium models, with application to a model of the Moroccan economy', *Review of Economics and Statistics* **80**(4), 520–534.

Amemiya, T. (1982), 'Two-stage least absolute deviations estimator', *Econometrica* **50**(3), 689–711.

Andrews, D. W. (1987), 'Consistency in nonlinear econometric models: A generic uniform law of large numbers', *Econometrica* **55**(6), 1465–1471.

Bahadur, R. and Savage, L. (1956), 'The nonexistence of certain statistical procedures in nonparametric problem', *Annals of Mathematical Statistics* **27**(4), 1115–1122.

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.

Barro, R. and Sala-i-Martin, X. (1991), 'Convergence across states and regions', *Brookings Papers on Economic Activity* **1991**(1), 107–182.

Bassett, G. and Koenker, R. (1978), 'Asymptotic theory of least absolute error regression', *Journal of the American Statitical Association* **73**(363), 618–622.

Bloomfield, P. and Steiger, W. L., eds (1983), *Least Absolute Deviations: Theory, Applications, and Algorithms*, Birkhauser, Boston.

Boldin, M. V., Simonova, G. I. and Tyurin, Y. N. (1997), *Sign-Based Methods in Linear Statistical Models*, Vol. 162 of *Translations of Mathematical Monographs*, American Mathematical Society, Maryland.

Buchinsky, M. (1995), 'Estimating the asymptotic covariance matrix for quantile regression models', *Journal of Econometrics* **68**(2), 303–338.

Buchinsky, M. (1998), 'Recent advances in quantile regression models: A practical guideline for empirical research', *Journal of Human Resources* **33**(1), 88–126.

Buehler, R. J. (1983), Fiducial inference, *in* (Johnson et al. 1983).

Chernozhukov, V., Hansen, C. and Jansson, M. (2009), 'Finite sample inference for quantile regression models', *Journal of Econometrics* **152**(2), 93–103.

Chernozhukov, V., Hong, H. and Tamer, E. (2007), 'Estimation and confidence regions for parameter sets in econometric models', *Econometrica* **75**(5), 1243–1284.

Coudin, E. and Dufour, J.-M. (2007), Finite-sample distribution-free inference in linear median regressions under heteroskedasticity and nonlinear dependence of unknown form, Technical report, McGill University, CIRANO and CIREQ, Montréal, and CREST, Paris.

Coudin, E. and Dufour, J.-M. (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.

Dodge, Y., ed. (1997), *L_1-Statistical Procedures and Related Topics*, number 31 *in* 'Lecture Notes - Monograph Series', Institute of Mathematical Statistics, Hayward, CA.

Dufour, J.-M. (1990), 'Exact tests and confidence sets in linear regressions with autocorrelated errors', *Econometrica* **58**(2), 475–494.

Dufour, J.-M. (1997), 'Some impossibility theorems in econometrics, with applications to structural and dynamic models', *Econometrica* **65**(6), 1365–1389.

Dufour, J.-M. (2003), 'Identification, weak instruments and statistical inference in econometrics. presidential address to the canadian economics association', *Canadian Journal of Economics* **36**(4), 767–808.

Dufour, J.-M. (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 Jasiak, J. (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., Jouneau, F. and Torrès, O. (2008), On (non)-testability : Applications to linear and nonlinear semiparametric and nonparametric regression models, Technical report, Department of Economics and CIREQ, McGill University, and GREMARS, Université de Lille 3.

Dufour, J.-M. and Kiviet, J. (1998), 'Exact inference methods for first-order autoregressive distributed lag models', *Econometrica* **82**(1), 79–104.

Dufour, J.-M. and Taamouti, M. (2005), 'Projection-based statistical inference in linear structural models with possibly weak instruments', *Econometrica* **73**(4), 1351–1365.

Dufour, J.-M. and Valéry, P. (2006), On a simple two-stage closed-form estimator for a stochastic volatility in a general linear regression, *in* T. B. Fomby and D. Terrell, eds, 'Advances in Econometrics, Volume 20: Econometric Analysis of Economic and Financial Time Series, in Honor of Clive Granger and Robert Engle, Part A', Elsevier Science, Oxford, U.K., pp. 259–288.

Dufour, J.-M. and Valéry, P. (2009), 'Exact and asymptotic tests for possibly non-regular hypotheses on stochastic volatility models', *Journal of Econometrics* **150**, 193–206.

Dwass, M. (1957), 'Modified randomization tests for nonparametric hypotheses', *Annals of Mathematical Statistics* **28**(1), 181–187.

Efron, B. (1998), 'R. A. Fisher in the 21st Century', *Statistical Science* **13**(2), 95–122.

El Bantli, F. and Hallin, M. (1999), 'L_1 estimation in linear models with heterogeneous white noise', *Statistics and Probability Letters* **45**, 305–315.

Fisher, R. A. (1930), 'Inverse probability', *Proc. Cambrige Philos. Soc.* **26**, 528–535.

Fitzenberger, B. (1997), 'The moving blocks bootstrap and robust inference for linear least squares and quantile regressions', *Journal of Econometrics* **82**(2), 235–287.

Gallant, A., Hsieh, D. and Tauchen, G. (1997), 'Estimation of stochastic volatility models with diagnostics', *Journal of Econometrics* **81**(1), 159–192.

Godambe, V. (2001), Estimation of median: Quasi-likelihood and optimum estimating functions, Discussion Paper 2001-04, Department of Statistics and Actuarial Sciences, University of Waterloo.

Hahn, J. (1997), 'Bayesian boostrap of the quantile regression estimator: A large sample study', *International Economic Review* **38**(4), 795–808.

Hannig, J. (2006), On fiducial inference - the good, the bad and the ugly, Technical report, Department of Statistics, Colorado State University, Fort Collins, Colorado.

Hansen, L. (1982), 'Large sample properties of generalized method of moments estimators', *Econometrica* **50**(4), 1029–1054.

Hodges, J. L. and Lehmann, E. L. (1963), 'Estimates of location based on rank tests', *The Annals of Mathematical Statistics* **34**(2), 598–611.

Hong, H. and Tamer, E. (2003), 'Inference in censored models with endogenous regressors', *Econometrica* **71**(3), 905–932.

Honore, B. E. and Hu, L. (2004), 'On the performance of some robust instrumental variables', *Journal of Business & Economic Statistics* **22**(1), 30–39.

Jaeckel, L. A. (1972), 'Estimating regression coefficients by minimizing the dispersion of the residuals', *The Annals of Mathematical Statistics* **40**(5), 1449–1458.

Johnson, N. L., Kotz, S. and Read, C., eds (1983), *Encyclopedia of Statistical Sciences, Volume 3*, John Wiley & Sons, New York.

Jureckova, J. (1971), 'Nonparametric estimate of regression coefficients.', *The Annals of Mathematical Statistics* **42**(4), 1328–1338.

Knight, K. (1989), 'Limit theory for auto-regressive-parameter estimates in an infinite-variance random walk', *Canadian Journal of Statistics* **17**(3), 261–278.

Knight, K. (1998), 'Limiting distributions for L_1 regression estimators under general conditions', *Annals of Statistics* **6**(2), 755–770.

Koenker, R. (2005), *Quantile Regression*, Vol. 38 of *Econometric Society Monographs*, Cambridge University Press, Cambridge, U.K.

Koenker, R. and Bassett, Jr., G. (1978), 'Regression quantiles', *Econometrica* **46**(1), 33–50.

Koul, H. (1971), 'Asymptotic behavior of a class of confidence regions based on ranks in regression.', *The Annals of Mathematical Statistics* **42**(2), 466–476.

Newey, W. K. and McFadden, D. (1994), Large sample estimation and hypothesis testing, *in* R. F. Engle and D. L. McFadden, eds, 'Handbook of Econometrics', Vol. 4, Elsevier, North Holland, chapter 36, pp. 2113–2245.

Neyman, J. (1941), 'Fiducial argument and the theory of confidence intervals', *Biometrika* **32**(2), 128–150.

Parzen, M., Wei, L. and Ying, Z. (1994), 'A resampling method based on pivotal estimating functions', *Biometrika* **81**, 341–350.

Phillips, P. (1991), 'A shortcut to LAD estimator asymptotics', *Econometric Theory* **7**(04), 450–463.

Pollard, D. (1991), 'Asymptotics for least absolute deviation regression estimators', *Econometric Theory* **7**(2), 186–199.

Portnoy, S. (1991), 'Asymptotic behavior of regression quantiles in non-stationary, dependent cases', *Journal of Multivariate Analysis* **38**, 100–113.

Powell, J. (1983), 'The asymptotic normality of two-stage least absolute deviations estimators', *Econometrica* **51**(5), 1569–1576.

Powell, J. (1984), 'Least absolute deviations estimation for the censored regression model', *Journal of Econometrics* **25**(3), 303–325.

Powell, J. L. (1986), 'Censored regression quantiles', *Journal of Econometrics* **32**, 143–155.

Schweder, T. and Hjort, N. L. (2002), 'Confidence and likelihood', *Scandinavian Journal of Statistics* **29**(2), 309–332.

So, B. S. (1994), 'A sharp Cramér-Rao type lower-bound for median-unbiased estimators', *Journal of the Korean Statistical Society* **23**(1).

Sung, N. K., Stangenhaus, G. and David, H. (1990), 'A Cramér-Rao analogue for median-unbiased estimators', *Trabajos de Estadistica* **5**, 83–94.

Weiss, A. (1990), 'Least absolute error estimation in the presence of serial correlation', *Journal of Econometrics* **44**(1), 127–158.

Weiss, A. (1991), 'Estimating nonlinear dynamic models using least absolute error estimation', *Econometric Theory* **7**(1), 46–68.

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

Xie, M.-g. and Singh, K. (2013), 'Confidence distribution, the frequentist distribution estimator of a parameter: A review', *International Statistical Review* **81**(1), 3–39. With discussions by David R. Cox, Bradley Efron, Donald A.S. Fraser, Emanuel Parzen, Christian P. Robert, Tore Schweder and Nils Lid Hjort, 40-68.

Zhao, Q. (2001), 'Asymptotically efficient median regression in the presence of heteroskedasticity of unknown form', *Econometric Theory* **17**(4), 765–784.