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

n° 2008-33

Hodges-Lehmann Sign-based Estimators and Generalized Confidence Distributions in Linear Median Regressions with Moment-free Heterogenous Errors and Dependence of Unknown Form

E. COUDIN[1] – J.-M. DUFOUR[2]

---

1. Institut National de la Statistique et des Etudes Economiques (INSEE), and Centre de Recherche en Economie et Statistique (CREST). *Mailing address* : CREST-Laboratoire de Microéconométrie, Timbre J390, 15 Boulevard Gabriel Péri, 92245 Malakoff Cedex, France. Mail : elise.coudin@insee.fr

2. William Dow Chair in Political Economy, Professor of Economics, McGill University, Centre Interuniversitaire de Recherche en Analyse des Organisations (CIRANO), and Centre Interuniversitaire de Recherche en Economie Quantitative (CIREQ). *Mailing address*: Department of Economics, McGill University, Leacock Building, Room 519, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. Tel : (1) 514 398 8879; Fax : (1) 514 398 4938. Mail : jean-marie.dufour@mcgill.ca
Web page : http://www.jeanmariedufour.com

Hodges-Lehmann Sign-based Estimators and Generalized Confidence Distributions in Linear Median Regressions with Moment-free Heterogenous Errors and Dependence of Unknown Form*

Elise COUDIN – Jean-Marie DUFOUR

First version : May 2004
This version : December 2008

Abstract

This paper develops sign-based estimation methods for the parameters of a median regression in finite samples. We introduce p-value functions that give the confidence one may have in a certain value of the parameter given the realization of the sample and sign-based estimators that are the values associated with the highest confidence (p-value). The sign-based estimators are thus obtained using the Hodges-Lehmann principle of test inversion. They are expected to present the same robustness properties than the test statistics they come from and can straightly be associated with the finite-sample-based inference procedure described in Coudin and Dufour (2007). We also show they are median unbiased (under symmetry and estimator unicity) and present equivariance features similar to the LAD estimator. Consistency under point identification and asymptotic normality are provided and hold under weaker assumptions than the LAD estimator. However, small sample behavior is our first interest. By a Monte Carlo study of bias and RMSE, we show sign-based estimators perform better than the LAD in very heteroskedastic settings.

Key words : sign-based methods; median regression; test inversion; Hodges-Lehmann estimators; confidence distributions; p-value function; least absolute deviations 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.

Résumé

Cet article propose des outils d'estimation et d'inférence dans le cadre d'une régression linéaire sur la médiane, valides à distance finie sans recourir à des hypothèses paramétriques sur la distribution des erreurs. Nous considérons la fonction p-value qui associe un degré de confiance à chaque valeur testée du paramètre étant donné la réalisation de l'échantillon. Nous calculons des fonctions p-value simulées à partir de tests de Monte Carlo simultanés, puis des versions projetées pour chaque composante individuelle du paramètre. Nous suivons ensuite le principe d'inversion de test de Hodges-Lehmann et proposons d'utiliser comme estimateur, la valeur du paramètre associée au plus haut degré de confiance (à la plus forte p-value). L'estimateur de signe hérite des propriétés de robustesse des statistiques dont il est issu et peut être associé à la procédure d'inférence à distance finie décrite dans Coudin et Dufour (2007). Il est aussi sans biais pour la médiane sous unicité et symétrie des erreurs, et partage les propriétés d'invariance de l'estimateur des moindres valeurs absolues (LAD). Il est enfin convergent et asymptotiquement normal sous des conditions plus faibles que l'estimateur LAD. En échantillon fini, les simulations suggèrent qu'il est plus performant en termes de biais et d'erreur quadratique moyenne pour des processus très hétérogènes.

Mots clés : méthodes de signes ; régression sur la médiane ; inversion de test ; estimateurs de Hodges-Lehmann ; distribution de confiance ; fonctions p-value ; estimateur LAD ; régressions quantiles ; tests de signe ; inférence simultanée ; test de Monte Carlo ; méthodes de projection ; non normalité ; hétéroscédasticité ; dépendance sérielle ; GARCH ; volatilité stochastique.
Classification Journal of Economic Literature : C13, C12, C14, C15.

* The authors thank Marine Carrasco, Frédéric Jouneau, Marc Hallin, Thierry Magnac, Bill McCausland, Benoit Perron, and Alain Trognon for useful comments and constructive discussions. Earlier versions of this paper were presented at the 2005 Econometric Society World Congress (London), the Econometric Society European Meeting 2007 (Budapest) and CREST (Paris). This work was supported by the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Alexander-von-Humboldt Foundation (Germany), the Institut de Finance Mathématique de Montréal (IFM2), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Canada Council for the Arts (Killam Fellowship), the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, the Fonds de recherche sur la société et la culture (Québec), and the Fonds de recherche sur la nature et les technologies (Québec).

Contents

List of Definitions, Propositions and Theorems

List of Tables

List of Figures

1. Introduction

Fisher's fiducial distributions and other fiducial inference arguments [Fisher (1930), Buehler (1983), Efron (1998)] are not commonly used in econometrics because they require pivotal test statistics with known distribution to be available. This condition is not fulfilled in general, especially in semi-parametric or non-parametric 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. In this paper, we consider inference and estimation of the parameter of a linear median regression under a weak conditional 0-median assumption. The errors may be heteroskedastic, nonlinearly dependent with possibly noncontinuous distributions. We notice that, for any given sample size, the sign transform enables one to construct test statistics with known nuisance parameter free distribution without requiring additional parametric restrictions. This remark enables us to construct fiducial inference tools adapted to multidimensional parameters. For this, we shall combine sign-based tests of simultaneous hypothesis such as presented in Coudin and Dufour (2007), with increasing level with projection techniques. First, we construct the realized p-value function, which yields the *degree of confidence* one may have in each possible value of the parameter. Second, the parameter value with the highest confidence ($i.e.$ the highest p-value) provides a Hodges-Lehmann sign-based estimator [Hodges and Lehmann (1963)]. For each component, a projected p-value function provides a graphical illustration of the inference summary.

Fisher introduced the fiducial probability as a frequentist competitor to Bayesian posterior probabilities. Ignored for a long time, fiducial inference has recently enjoyed a renewed interest in the statistical literature with the introduction of confidence distributions and similar inference methods [Hannig (2006) for a review]. The confidence distribution is defined in the one-dimensional model as a distribution whose quantiles span all the possible confidence intervals [Schweder and Hjort (2002)]. The latter authors introduced it as a Neymanian interpretation of Fisher's fiducial distribution. This tool summarizes all the inference results on the parameter and gives a graphical representation of it. The confidence distribution is related to the p-values for testing hypotheses of the form $H_0(\beta_0): \ \beta = \beta_0$. The realized p-value can be seen as the *degree of confidence* one may have on the tested value. These tools can be constructed whenever a pivotal increasing function of the parameter with known distribution is available. The sign transform enables one to construct such pivots without imposing parametric restrictions on the sample. However, the sign-based statistics so constructed are discrete and only approximate confidence distributions or realized p-value functions can be obtained.

Then, we derive estimators and study their properties. Hodges and Lehmann (1963) proposed a general principle to directly derive estimators from test statistics for a given sample size. They suggest to invert a test for $H_0(\beta_0) : \beta = \beta_0$, and to choose the value of β which is "least rejected" by the test. First applied to the Wilcoxon's signed rank-statistic for estimating a shift or a location, this principle was adapted for a regression context by Jureckova (1971), Jaeckel (1972) and Koul (1971). The latter authors derived so-called R-estimators from rank or signed-rank statistics. In a multidimensional context, this leads one to select the value of β with the highest *degree of confidence* $i.e.$ with the highest p-value. The sign-based estimators are obtained by using sign-based tests. They inherit some of the attractive properties the tests they come from (robustness to model specification, gross errors and heteroskedasticity). We shall see that they alternatively can be computed by minimizing quadratic forms of the constrained signs (with probability one). So they have a classical GMM form [Hansen (1982), and Honore and Hu (2004) for GMM statistics involving signs]. We show that sign-based estimators are consistent and asymptotically normal under regularity conditions weaker than the ones required by the LAD estimator usual theory [Powell (1984), Weiss (1991), Fitzenberger (1997)].[1] In particular, asymptotic normality and consistency hold for heavy-tailed disturbances that may not possess finite variance. This interesting property is entailed by the sign transformation. Signs of residuals always possess finite moments so no further restriction on the disturbance moments is required to complete the proofs. In finite samples also, LAD and sign methods exhibit very different features. The simulation studies of bias and root mean squared error (RMSE) we present show that sign-based estimators are more robust than the LAD estimator in the presence of heteroskedasticity. The class of sign-based estimators includes some special cases studied in the statistical literature: Boldin, Simonova, and Tyurin (1997) derived *sign-estimators* from locally most powerful test statistics for $i.i.d.$ observations and fixed regressors. Instrumental versions of sign-based estimators are presented in Honore and Hu (2004) and Hong and Tamer (2003). Honore and Hu (2004) derived the so-called *median-based estimator* as an instrumental GMM version of the quantile estimator. The authors motivated to use the latter along with

[1]Other notable research on LAD estimators and their variants includes: the efficient weighted LAD of Zhao (2001), the smoothed LAD of Horowitz (1998), adaptations to allow for endogeneity [Amemiya (1982), Powell (1983), Hong and Tamer (2003)], nonlinear functional forms [Weiss (1991)] and generalization to quantile regressions [Koenker and Bassett (1978)].

other rank-based estimators for their general robustness properties. However, the major advantage of signs upon ranks is to easily deal with heteroskedastic disturbances. In the present paper, we do not assume $i.i.d.$ disturbances. We derive various sign-based statistics and associated sign-based estimators depending on the setup. Many heteroskedastic and possibly dependent schemes are covered and, when needed, an heteroskedasticity and autocorrelation correction is included in the estimator criterion function. Restricting on $i.i.d.$ cases, Honore and Hu (2004) observed in simulations that inference based on rank-based estimators performed better than the median-based one. In particular, the estimates of the asymptotic standard errors of the median-based estimator, that they obtained by kernel, were too small and the associated inference suffered from overrejection of the null hypothesis. Deriving sign-based estimators as Hodges-Lehmann estimators motivates us to definitely combine them to the finite sample inference method they come from, which is developed in Coudin and Dufour (2007). The latter, based on the exact distribution of the corresponding sign-based test statistics does not depend on any nuisance parameter and does control test levels in finite samples under heteroskedasticity and nonlinear dependence of unknown form. It combines Monte Carlo techniques [Dwass (1957), Barnard (1963) and Dufour (2006)], inversion and projection techniques [Dufour (1990, 1997), Dufour and Kiviet (1998), Abdelkhalek and Dufour (1998), Dufour and Jasiak (2001), Dufour and Taamouti (2005)].[2] It does not require one to estimate the error density at zero in contrast to tests based on kernel estimates of the LAD asymptotic covariance matrix.[3] Therefore, when the test criteria are modified to cover linear dependence, the resulting inference is asymptotically valid.

Finally, sign-based tests, projection-based confidence regions, projection-based p-values and sign-based estimators constitute a whole system of inference valid for any given sample size under very weak distributional assumptions and asymptotically valid for linear dependent errors.

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 introduce the sign-based estimators, which are obtained by maximizing the p-value function. Finite-sample properties of sign-based estimators are

[2] For 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 (2008).

[3] 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)] and resampling methods [Parzen, Wei, and Ying (1994)].

established in section 5 and asymptotic properties in section 6. In section 7, we present a simulation study of bias and RMSE. In section 8, we apply sign-based estimation for deriving robust estimates in two cases: first, in a financial setup involving large heteroskedasticity (S.&P. index); second, in a cross-sectional regional data set where the sample size is necessarily small (β-convergence of output levels across U.S. States). Section 9 concludes. Appendix A contains the proofs.

2. Framework

2.1. Model

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

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

where y_t is a dependent variable, $x_t = (x_{t1}, \ldots, x_{tp})'$ 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, $y = (y_1, \ldots, y_n)' \in \mathbb{R}^n$ will denote the dependent vector, $X = [x_1, \ldots, x_n]'$ the $n \times p$ matrix of explanatory variables, and $u = (u_1, \ldots, u_n)' \in \mathbb{R}^n$ the disturbance vector. Moreover, $F_t(\cdot | x_1, \ldots, x_n)$ represents the distribution function of u_t conditional on X. This framework is also used in Coudin and Dufour (2007).

The traditional form of a median regression assumes u_t's are $i.i.d.$ with median zero

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

Here, we relax the assumption that the u_t are $i.i.d.$ and consider instead 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 $s(u)$, the n-vector composed by the signs of its components. We assume the following assumption holds.

Assumption 2.1 SIGN MOMENT CONDITION. $E[s(u_t)x_{kt}] = 0$, *for* $k = 1, \ldots, p$, $t = 1, \ldots, n$, *and* $n \in \mathbb{N}$.

Assumption 2.1 is fulfilled if the disturbances are $i.i.d.$. Now let us introduce *adapted sequences* $\mathcal{S}(\mathbf{v}, \mathcal{F}) = \{v_t, \mathcal{F}_t : t = 1, 2, \ldots\}$ where v_t is any measurable function of $W_t = (y_t, x_t')'$, $\mathcal{F}_t$ is a σ-field in Ω, $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s < t$, $\sigma(W_1, \ldots, W_t) \subset \mathcal{F}_t$ and $\sigma(W_1, \ldots, W_t)$

is the σ-algebra spanned by $W_1, \ldots, W_t$. Assumption 2.1 is also fulfilled if the signs satisfy a martingale difference with respect to the past information $\mathcal{F}_t = \sigma(W_1, \ldots, W_t)$:

$$E[s(u_t)|\mathcal{F}_{t-1}] = 0, \ \ \forall t \geq 1. \tag{2.3}$$

Assumption 2.1 covers many weakly dependent processes including usual linear dependent processes, such as $AR(1)$ disturbances with normal innovations and mean zero. This has been pointed out by Fitzenberger (1997). Assumption 2.1 also holds when $\mathbf{u}$ satisfies the conditional mediangale condition defined in Coudin and Dufour (2007), $i.e.$ when $\{s(u_t) : t = 1, 2, \ldots\}$ is a martingale difference with respect to $\{\mathcal{F}_t = \sigma(W_1, \ldots, W_t, X)\}$:

Assumption 2.2 WEAK CONDITIONAL MEDIANGALE. *Let* $\mathcal{F}_t = \sigma(u_1, \ldots, u_t, X)$, *for* $t \geq 1$. $\mathbf{u}$ *in the adapted sequence* $\mathcal{S}(\mathbf{u}, \mathcal{F})$ *is a strict mediangale conditional on* X *with respect to* $\{\mathcal{F}_t : t = 1, 2, \ldots\}$ *iff* $\mathsf{P}[u_1 < 0|X] = \mathsf{P}[u_1 > 0|X]$ *and*

$$\mathsf{P}[u_t < 0|u_1, \ldots, u_{t-1}, X] = \mathsf{P}[u_t > 0|u_1, \ldots, u_{t-1}, X], \ \textit{for } t > 1. \tag{2.4}$$

This setup allows for discrete distributions with mass at zero. When the distributions of u_t possess a mass at zero, the sign operator is redefined as $\tilde{s}(a, V) = s(a) + \left[1 - s(a)^2\right] s(V - 0.5)$, where $V \sim \mathcal{U}(0, 1)$ and is independent of a. With no mass at zero and no matrix X, this mediangale concept coincides with the one defined in Linton and Whang (2007) together with other quantilegales.[4] Assumption 2.1 is exploited to construct test statistics.

2.2. Sign-based statistics and Monte Carlo tests

For testing $H_0(\beta_0) : \ \beta = \beta_0$ vs. $H_1(\beta_0) : \ \beta \neq \beta_0$ in model (2.1), we consider general quadratic forms involving the vector of the residual signs for the constrained model $s(y - X\beta_0)$:

$$D_S(\beta_0, \Omega_n) = s(y - X\beta_0)' X \Omega_n(s(y - X\beta_0), X) X' s(y - X\beta_0) \tag{2.5}$$

where $\Omega_n(s(y - X\beta_0), X)$ is a $p \times p$ positive definite weight matrix that may depend on the constrained signs. In Coudin and Dufour (2007), we developed distribution-free Monte Carlo tests. We briefly summarize it. If the disturbances satisfy the mediangale Assumption 2.2, the sign-based statistics satisfying equation (2.5) are shown to be pivotal functions under $H_0(\beta_0)$. The distribution of the statistic conditional on the realization of X, is perfectly specified and can be simulated. Monte Carlo tests with controlled levels are constructed

[4]Linton and Whang (2007) define that u_t is a mediangale if $E(\psi_{\frac{1}{2}}(u_t)|\mathcal{F}_{t-1}) = 0, \ \forall t$, where $\mathcal{F}_{t-1} = \sigma(u_{t-1}, u_{t-2}, \ldots)$ and $\psi_{\frac{1}{2}}(x) = \frac{1}{2} - \mathbf{1}_{(-\infty,0)}(x)$. The specification of the sign function which does not make difference between a positive and a null number is clearly adapted to continuous distributions.

in the following way. For testing $H_0(\beta_0)$ vs. $H_1(\beta_0)$ with level $\alpha \in [0,1]$, we denote $D_S^{(0)} = D_S(\beta_0)$ the observed statistics, $(D_S^{(1)}, \ldots, D_S^{(N)})'$ an N-vector of independent replicates drawn from the same distribution as $D_S(\beta_0)$ and $(W^{(0)}, \ldots, W^{(N)})'$, a $N+1$-vector of $i.i.d.$ uniform variables. A Monte Carlo test for $H_0(\beta_0)$ consists in rejecting the null hypothesis whenever the empirical p-value $\tilde{p}_N^{D_S}(\beta_0)$ is smaller than α, where

$$\tilde{p}_N^{D_S}(x) = \frac{N\tilde{G}_N(x) + 1}{N+1} \tag{2.6}$$

and $\tilde{G}_N(x) = 1 - \frac{1}{N}\sum_{i=1}^{N} s_+(x - T^{(i)}) + \frac{1}{N}\sum_{i=1}^{N} \delta(T^{(i)} - x)s_+(W^{(i)} - W^{(0)}),$ with $s_+(x) = \mathbf{1}_{[0,\infty)}(x)$, $\delta(x) = \mathbf{1}_{\{0\}}$. The empirical p-value is based on a randomized tie-breaking procedure which allows one to control the level when the statistics are discrete. When the number of replicates N is such that $\alpha(N+1)$ is an integer, the level of the Monte Carlo test is equal to α for any sample size n [see Dufour (2006)]. Next, simultaneous confidence regions for the entire parameter β are obtained by inverting those simultaneous tests. The simultaneous confidence region $C_{1-\alpha}(\beta)$

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

which contains all the values β^* with empirical p-value $\tilde{p}_N^{D_S}(\beta^*)$ [associated with the test of $H_0(\beta^*):\ \beta = \beta^*$] 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, 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 (2007).

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*

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

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

$(-\infty,\ CD^{-1}(\alpha)]$ constitutes a one-sided stochastic confidence interval with coverage probability α,[5] and the realized confidence $CD(\beta_0; y; x)$ is the p-value of the one-sided hypothesis $H_0^*(\beta_0)\ :\ \ \beta \leq \beta_0$ 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 $H_1(\beta_0)\ :\ \ \beta \neq \beta_0$ 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 "$\beta \leq \beta_0$" 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 $H_0^*(\beta_0) : \beta \leq \beta_0$ v.s. $H_1^*(\beta_0) : \beta > \beta_0$.* 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 distribution function F that is independent of β and free of any nuisance parameter), $F\big(T(\beta)\big)$ is uniformly distributed

[5] for continuous distributions, just note that $P_\beta[\beta \leq CD^{-1}(\alpha)] = P_\beta\{CD(\beta) \leq CD\big(CD^{-1}(\alpha)\big)\} = P_\beta\{CD(\beta) \leq \alpha]\} = \alpha$

and satisfies conditions for providing a confidence distribution. Let $\hat{\beta}$ 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 $\hat{\beta}^{obs}$ is large, with p-value $P_{\beta_0}[\beta > \hat{\beta}^{obs}]$. Then,

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

where $F_{\beta_0}(\hat{\beta})$ is the sampling distribution of $\hat{\beta}$. 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 $\hat{\beta}$ obtained by repeated samplings whereas the confidence distribution is an ex-post object which contains the confidence statements one can have on the value of β given $y, x, \hat{\beta}^{obs}$.

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

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

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

$$CD(\beta_0) = P_{\beta_0}[\beta > \hat{\beta}^{obs}] + P[U_{\hat{\beta}}^{(0)} > U_\beta] P_{\beta_0}[\beta = \hat{\beta}^{obs}]. \tag{3.4}$$

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

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

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

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

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 $H_0^*(\beta_0) : \beta \leq .1$ at

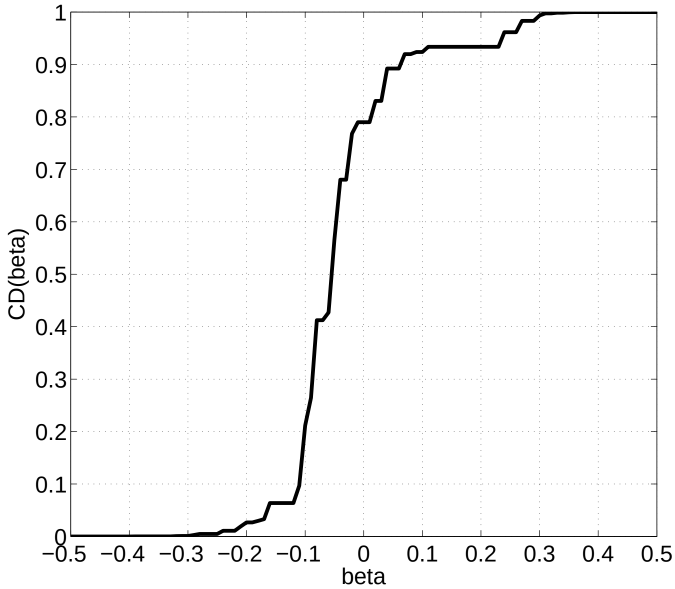


Figure 1. Simulated confidence distribution cumulative function based on SST.

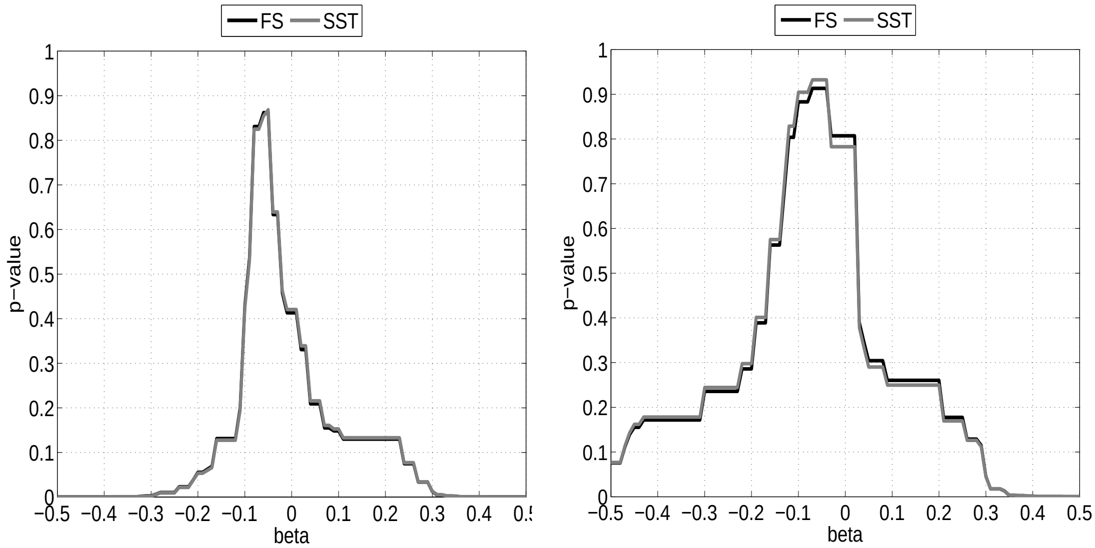
10% can be done by reading $CD(.1)$, which equals the p-value of H_0, here .88. The test accepts H_0^*. Further, $(-\infty, .15]$ constitutes a one-sided confidence interval for β with level $.95$.

Realized p-value functions for discrete statistics. Another interesting object is the realized p-value 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 (2.6). Figure 2 compares 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, $[-.12, .14]$ 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 *parameter identification*. When the p-values are low (or high) whatever the value of β, 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



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

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

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 followings, $y_i = \beta_2 x_i + u_i,\ i = n_1 + 1, \ldots, n_1 + n_2, (u_i, x_i) \overset{iid}{\sim} \mathcal{N}(0, I_2)$, 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 which we can interpret as a lack of identification: 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 (2.5). Simulated p-values lead to a mapping for which we have a 3-dimensional representation for $p = 2$. Consider the model: $y_i = \beta^1 x_{1i} + \beta^2 x_{2i} + u_i,\ i = 1, \ldots, n, (u_i, x_{1i}, x_{2i}) \overset{iid}{\sim} \mathcal{N}(0, I_3)$, $\beta = (\beta^1, \beta^2) = (0, 0)'$, $y = (y_1, \ldots, y_n)'$, $u = (u_1, \ldots, u_n)'$, $x_1 = (x_{11}, \ldots, x_{1n})'$, $x_2 = (x_{21}, \ldots, x_{2n})'$ 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

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

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

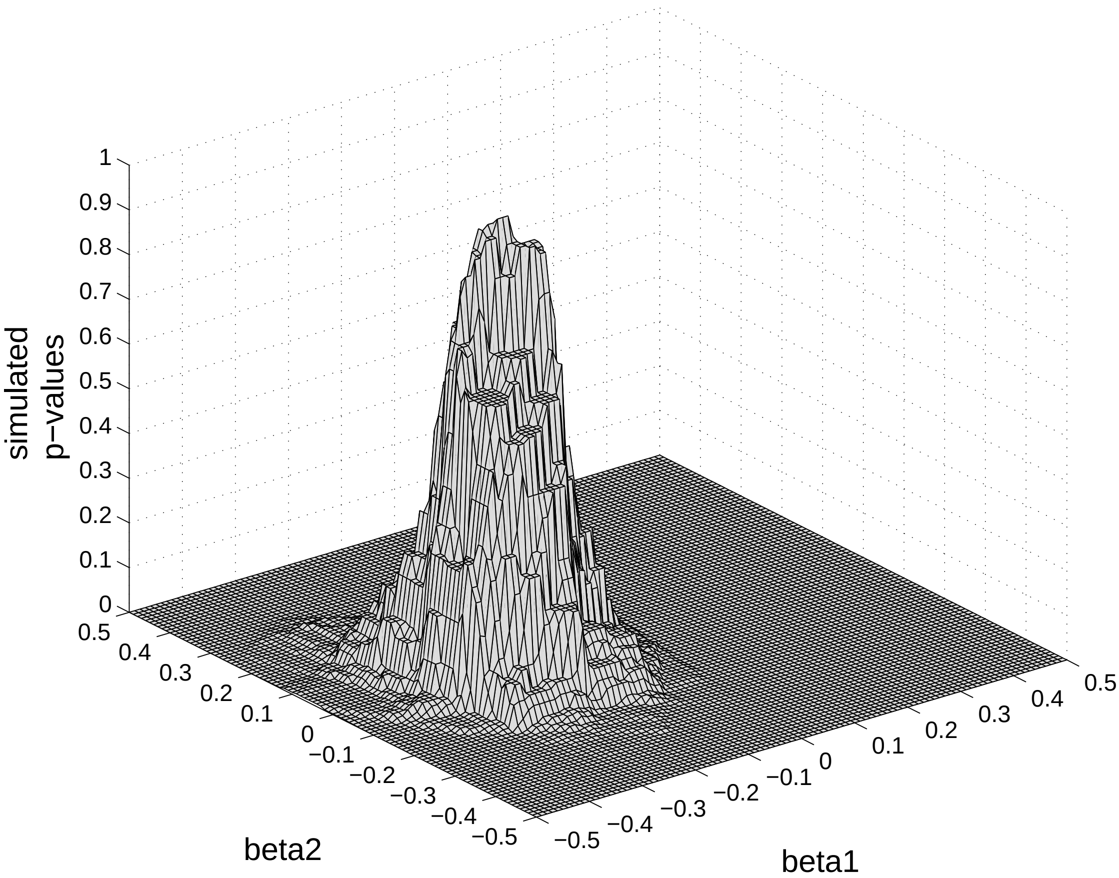
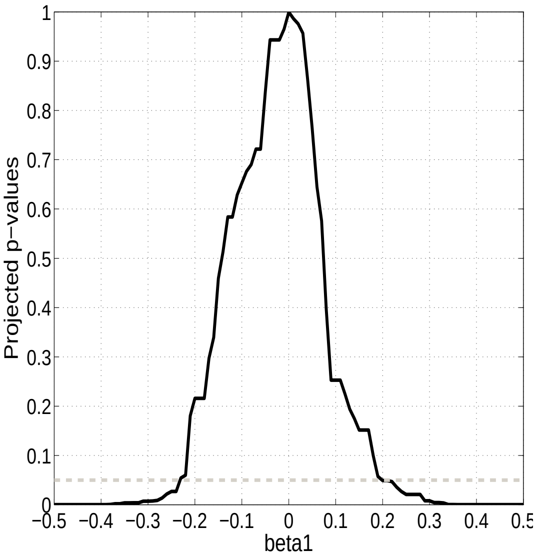


Figure 3. Simulated p-value functions based on SF ($n = 200$, $N = 9999$).

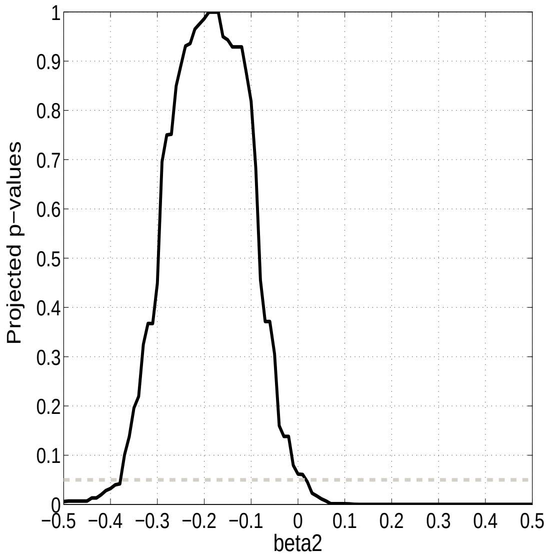
dimensions ($p > 2$), a complete graphical representation is not available anymore. However, one can consider projection-based realized p-value functions for each individual component of the parameter of interest in a similar way than projection-based confidence intervals. For this, we apply the general 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:

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

Figure 4 presents projection-based confidence intervals for the individual parameters of the previous 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 $.06$.



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



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

Figure 4. Projection-based p-values.

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 the 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 optimization problems

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

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

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

Then, the following proposition holds.

Proposition 4.1 *Under Assumption 4.1, $M1$ and $M2$ are equal with probability one.*

Maximizing $\mathrm{p}(\beta)$ is equivalent (in probability) to minimizing $T(\beta)$ if Assumption 4.1 holds. Under the mediangale Assumption 2.2, any sign-based statistic D_S does satisfy Assumption 4.1. Consequently,

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

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.

The whole argmin set of (4.4) remains informative in models with sets of observationally equivalent values of β [see Chernozhukov, Tamer, and Hong (2006)]. The identified feature of those models is a set instead of a point value. Any inference approach relying on the consistency of a point estimator (which assumes point identification), gives misleading results, but the estimation of the whole set can be exploited. Let us remind that the Monte Carlo sign-based inference method [Coudin and Dufour (2007)] does not rely on identification conditions and leads to valid results in any case.

The sign-based estimators studied by Boldin, Simonova, and Tyurin (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

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

The weighted sign-based estimators are solutions of

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

where $\tilde{X} = diag(d_1, \ldots, d_n)X$ 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

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

where $\tilde{X} = diag\big(f_1(0|X), \ldots, f_n(0|X)\big)X$ 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 2.1 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 set. 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 connex. To see that it is a union of convex sets just remark that the reciprocal image of n fixed signs is convex.

5.1. Invariance

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. *If* $\hat{\beta}(Y, X) \in M_2(Y, X, D_S^{\Omega_n})$, $i.e.$ *is a solution of (4.4),* *then*

$$\lambda\hat{\beta}(Y,X) \in M_2(\lambda Y, X, D_S^{\Omega_n}), \quad \forall\lambda \in \mathbb{R} \tag{5.12}$$

$$\hat{\beta}(Y,X)+\gamma \in M_2(Y+X\gamma, X, D_S^{\Omega_n}), \quad \forall\gamma \in \mathbb{R}^p \tag{5.13}$$

$$A^{-1}\hat{\beta}(Y,X) \in M_2(Y, XA, D_S^{\Omega_n}), \quad \textit{for any nonsingular } k\times k \textit{ matrix } A. \tag{5.14}$$

To prove this property, it is sufficient to write down the different optimization problems. Equation (5.12) states 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. Equation (5.13) states a form of location invariance, while (5.14) states a reparameterization invariance with respect to the design matrix: the transformation on $\hat{\beta}$ is given by the inverse of the reparameterization scheme.

5.2. Median unbiasedness

Moreover, 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 $\hat{\beta}$ is the unique solution of minimization problem* (4.4)*, then $\hat{\beta}$ is median unbiased, that is,*

$$Med(\hat{\beta}-\beta_0)=0$$

where β_0 is the true value.

6. Asymptotic properties

We demonstrate consistency 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. Their finite-sample behavior also presents useful features. Finally, sign-based estimators are asymptotically normal.

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:

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 $E|x_{ht}|^{r+1} < \Delta < \infty,\ h = 1, \ldots, p,\ t = 1, \ldots, n,\ \forall n \in \mathbb{N}$.*

Assumption 6.3 COMPACTNESS. *$\beta \in Int(\Theta)$, 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 $n \in \mathbb{N}$,*
$$P[f_t(0|X) > f_L] > p_1,\ \forall t = 1, \ldots, n,\ \textit{a.s.}$$
2. *$f_t(.|X)$ 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*
$$\liminf_{n\to\infty} \frac{1}{n} \sum_t P[|x_t'\delta| > \tau | f_t(0|x_1, \ldots, x_n) > f_L] > 0.$$

Assumption 6.6 UNIFORMLY POSITIVE DEFINITE WEIGHT MATRIX. *$\Omega_n(\beta)$ is symmetric definite positive for all β in Θ.*

Assumption 6.7 LOCALLY POSITIVE DEFINITE WEIGHT MATRIX NEAR β_0. *$\Omega_n(\beta)$ is symmetric definite positive for all β in a neighborhood of β_0.*

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

Theorem 6.1 CONSISTENCY. *Under model (2.1) with the Assumptions 2.1, 6.1-6.6, any sign-based estimator of the type,*
$$\hat{\beta}_n(\Omega_n) \in \operatorname*{argmin}_{\beta \in \mathbb{R}^p} s'(Y - X\beta) X \Omega_n\big(s(y - X\beta), X\big) X' s(Y - X\beta), \tag{6.15}$$
or
$$\hat{\beta}_n^{2S}(\Omega_n) \in \operatorname*{argmin}_{\beta \in \mathbb{R}^p} s'(Y - X\beta) X \hat{\Omega}_n\big(s(y - X\hat{\beta}), X\big) X' s(Y - X\beta), \tag{6.16}$$
where $\hat{\beta}$ stands for any (first step) consistent estimator of β, is consistent. $\hat{\beta}_n^{2S}$ defined in equation (6.16) is still consistent if Assumption 6.6 is replaced by Assumption 6.7.

Let us interpret precisely Assumptions 6.1-6.7 and compare them to the ones required for LAD and quantile estimator consistency [see Fitzenberger (1997) and Weiss (1991) for the most general setups]. Assumptions on mixing (6.1), compactness (6.3) and point identification (6.4, 6.5, 6.6) are classical. The mixing setup 6.1 is needed to apply a generic weak law of large numbers [see Andrews (1987) and White (2001)]. It was used by Fitzenberger (1997) to show LAD and quantile estimator consistency with stationary linearly dependent processes. It covers, among other processes, stationary ARMA disturbances with continuously distributed innovations. Point identification is provided by Assumptions 6.5 and 6.4. Assumption 6.5 is similar to Condition ID in Weiss (1991). Assumption 6.4 is usual in the LAD estimator asymptotics.[6] It is analogous to Fitzenberger (1997)'s conditions (ii.b and c) and Weiss (1991)'s 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. So-called *diffusivity*, $\frac{1}{2f(0)}$, can indeed be seen as an alternative measure of dispersion adapted to median-unbiased estimators. It measures the vertical spread of a density rather than its horizontal spread and is involved in Cramér-Rao-type lower bound for median-unbiased estimators [see Sung, Stangenhaus, and David (1990) and So (1994)]. Besides, in Assumptions 6.6 and 6.7, the weight matrix Ω_n is supposed to be invertible for estimators obtained in one step whereas only a local invertibility is needed for two-step sign-estimators. One difference with the LAD asymptotic properties relies on Assumption 6.2. For sign consistency, only the second-order moments of x_t have to be finite, which differs from Fitzenberger (1997) who supposed the existence of at least third-order moments. And above all, we do not assume the existence of second-order moments on the disturbances u_t. The disturbances indeed appear in the objective function only through their sign transforms which possess finite moments up to any order. Consequently, no additional restriction should be imposed on the disturbance process (in addition to regularity conditions on the density). Those points will entail a more general CLT than the one stated for the LAD/quantile estimators in Fitzenberger (1997) and Weiss (1991).

6.2. Asymptotic normality

Sign-based estimators are asymptotically normal. This also holds under weaker assumptions than the ones needed for LAD estimator asymptotic normality. Sign-based estimators are specially adapted for heavy-tailed disturbances that may not possess finite variance. The assumptions we need are the following ones.

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

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

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

Under the conditions 2.1, 6.1, 6.2 and 6.8, we can define $L(\beta)$, the derivative of the limiting objective function at β:

$$L(\beta) = \lim_{n\to\infty} \frac{1}{n} \sum_t E\big[x_t x_t' f_t\big(x_t'(\beta - \beta_0)|x_1, \ldots, x_n\big)\big] = \lim_{n\to\infty} L_n(\beta). \tag{6.17}$$

where

$$L_n(\beta) = \frac{1}{n} \sum_t E\big[x_t x_t' f_t\big(x_t'(\beta - \beta_0)|x_1, \ldots, x_n\big)\big]. \tag{6.18}$$

The other assumptions are merely used to show asymptotic normality.

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

Assumption 6.10 DEFINITE POSITIVENESS OF L_n. *$L_n(\beta_0)$ is positive definite uniformly in n.*

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

Then, we have the following result.

Theorem 6.2 ASYMPTOTIC NORMALITY. *Under the conditions for consistency (2.1, 6.1-6.6), and 6.9-6.11, we have:*

$$S_n^{-1/2} \sqrt{n}(\hat{\beta}_n(\varOmega_n) - \beta_0) \xrightarrow{d} N(0, I_p) \tag{6.19}$$

where

$$S_n = [L_n(\beta_0)\varOmega_n L_n(\beta_0)]^{-1} L_n(\beta_0)\varOmega_n J_n \varOmega_n L_n(\beta_0)[L_n(\beta_0)\varOmega_n L_n(\beta_0)]^{-1}$$

and

$$L_n(\beta_0) = \frac{1}{n} \sum_t E\big[x_t x_t' f_t\big(0|x_1, \ldots, x_n\big)\big]. \tag{6.20}$$

Remark that when $\varOmega_n = \hat{J}_n^{-1}$, we have

$$[L_n(\beta_0)\hat{J}_n^{-1} L_n(\beta_0)]^{-1/2} \sqrt{n}(\hat{\beta}_n(\hat{J}_n^{-1}) - \beta_0) \xrightarrow{d} \mathrm{N}(0, I_p). \tag{6.21}$$

This corresponds to the use of optimal instruments and quasi-efficient estimation. $\hat{\beta}(\hat{J}_n^{-1})$ 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.[7] 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 2.2], both $\hat{\beta}_n((X'X)^{-1})$ and $\beta(\hat{J}_n^{-1})$ are asymptotically normal with variance

$$S_n = \ [L_n(\beta_0)]^{-1} E\left[(1/n) \sum_{t=1}^{n} x_t x_t'\right] [L_n(\beta_0)]^{-1}.$$

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

$$S_n = \frac{1}{4f(0)^2} E(x_t x_t')^{-1}. \tag{6.22}$$

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, Simonova, and Tyurin (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.

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

6.3. Asymptotic or projection-based sign-confidence intervals?

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 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 (2007)- see subsection 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 2.2 holds. When only the sign moment condition 2.1 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 (2007)]. 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 2.1 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. 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. 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:

$$y_t = x_t'\beta_0 + u_t, \tag{7.1}$$

where $x_t = (1, x_{2,t}, x_{3,t})'$ and β_0 are 3×1 vectors. We denote the sample size n. Monte Carlo studies are based on M 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 neither consistent anymore. 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 presents a very debalanced 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 or arbitrary 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 debalanced 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

Table 1. Simulated models.

A1:	Normal HOM:	$(x_{2,t}, x_{3,t}, u_t)' \overset{i.i.d}{\sim} \mathcal{N}(0, I_3),\ t = 1, \ldots, n$
A2:	Normal HET:	$(x_{2,t}, x_{3,t}, \tilde{u}_t)' \overset{i.i.d}{\sim} \mathcal{N}(0, I_3)$ $u_t = min\{3, max[0.21, \lvert x_{2,t}\rvert]\} \times \tilde{u}_t,\ t = 1, \ldots, n$
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:	Deb. design mat.:	$x_{2,t} \sim \mathcal{B}(1, 0.3),\ x_{3,t} \overset{i.i.d.}{\sim} \mathcal{N}(0, .01^2),$ $u_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$, x_t, u_t independent, $t = 1, \ldots, n$.
B5:	Cauchy dist.:	$(x_{2,t}, x_{3,t})' \sim \mathcal{N}(0, I_2),$ $u_t \overset{i.i.d.}{\sim} \mathcal{C}$,$x_t, u_t,$ independent, $t = 1, \ldots, n$.
B6:	Stoc. Volat.:	$(x_{2,t}, x_{3,t})' \overset{i.i.d.}{\sim} \mathcal{N}(0, I_2), u_t = exp(w_t/2)\epsilon_t$ with $w_t = 0.5 w_{t-1} + v_t$, where $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$, $v_t \overset{i.i.d.}{\sim} \chi_2(3)$, $x_t, u_t,$ independent, $t = 1, \ldots, n$.
B7:	Nonstat. GARCH(1,1):	$(x_{2,t}, x_{3,t}, \epsilon_t)' \overset{i.i.d.}{\sim} \mathcal{N}(0, I_3),\ t = 1, \ldots, n,$ $u_t = \sigma_t \epsilon_t,\ \sigma_t^2 = 0.8 u_{t-1}^2 + 0.8 \sigma_{t-1}^2.$
B8:	Exp. Var.:	$(x_{2,t}, x_{3,t}, \epsilon_t)' \overset{i.i.d.}{\sim} \mathcal{N}(0, I_3), u_t = exp(.2t)\epsilon_t.$
C9:	AR(1)-HOM, $\rho_u = .5$:	$(x_{2,t}, x_{3,t}, \nu_t^u)' \sim \mathcal{N}(0, I_3), t = 2, \ldots, n,$ $u_t = \rho_u u_{t-1} + \nu_t^u,$ $(x_{2,1}, x_{3,1})' \sim \mathcal{N}(0, I_2)$, ν_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, $\rho_u = .9$:	$(x_{2,t}, x_{3,t}, \nu_t^u)' \sim \mathcal{N}(0, I_3), t = 2, \ldots, n,$ $u_t = \rho_u u_{t-1} + \nu_t^u,$ $(x_{2,1}, x_{3,1})' \sim \mathcal{N}(0, I_2)$, ν_1^u insures stationarity.

Table 2. Simulated bias and RMSE.

$n = 50$		OLS		LAD		SF		2SSHAC	
$S = 1000$		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
	$\|\|\boldsymbol{\beta}\|\|$*	**.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
	$\|\|\boldsymbol{\beta}\|\|$	**.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
	$\|\|\boldsymbol{\beta}\|\|$	**.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
	$\|\|\boldsymbol{\beta}\|\|$	**.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
	$\|\|\boldsymbol{\beta}\|\|$	**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
	$\|\|\boldsymbol{\beta}\|\|$	**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
	$\|\|\boldsymbol{\beta}\|\|$	**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
	$\|\|\boldsymbol{\beta}\|\|$	$\mathbf{> 10^{10}}$	$\mathbf{> 10^{10}}$	$\mathbf{> 10^{10}}$	$\mathbf{> 10^{10}}$	**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
	$\|\|\boldsymbol{\beta}\|\|$	**.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
	$\|\|\boldsymbol{\beta}\|\|$	**.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
	$\|\|\boldsymbol{\beta}\|\|$	**.070**	**1.29**	**.027**	**.430**	**.059**	**1.36**	**.054**	**1.37**

* $\|\|.\|\|$ stands for the euclidian norm.

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 debalanced 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 (2007) 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. Illustrations

In this section, we go back to the two illustrations presented in Coudin and Dufour (2007) 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.

8.1. Drift estimation with stochastic volatility in the error term

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 Valéry and Dufour (2004) 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,

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

and we allow that $\{u_t : t = 1, \ldots, 16127\}$ exhibits stochastic volatility or nonlinear heteroskedasticity of unknown form. White and Breush-Pagan tests for heteroskedasticity both reject homoskedasticity at 1%.[8]

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 heteroskedasticy can bias the subsample estimates. The Wall Street krach occurred between October, 24th (*Black Thursday*) and October, 29th (*Black Tuesday*). Hence, the second subsample corresponds to the period just before the krach (September), the krach period (October) and the early beginning of the Great Depression (November and December). Heteroskedasticity tests reject homoskedasticity for both subsamples.[9]

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 $SHAC$);[10] 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, we note that the OLS estimates are importantly biased and are greatly unreliable in the presence of heteroskedasticity. Hence, they are just reported for comparison sake. 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 function evaluated at

[8] See Coudin and Dufour (2007): White: 499 (p-value=.000) ; BP: 2781 (p-value=.000).

[9] 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.

[10] see Coudin and Dufour (2007)

Table 3. Constant and drift estimates.

	Whole sample	Subsamples	
Constant parameter (a)	*(16120 obs)*	*1929 (291 obs)*	*1929 (90 obs)*
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.

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 β-convergence across US 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 between 1880 and 1988. They used nonlinear least squares to estimate equations of the form

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

$i = 1, \ldots, 48,\ T = 8,\ 10$ 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

Table 4. Summary of regression diagnostics.

Period	Heterosked.*		Nonnormality**		Influent. obs.**		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 Breush-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, studendized or standardized residuals > 3, DFbeta and Cooks distance have been performed and lead to suspicions for nonnormality, outlier or high influential observation presence.

are probably combined. We wonder which part of that variability is really due to business cycles and which part is only due to the nonrobustness 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 equation:

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

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

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

Period	*Basic equation*		*Equation with regional dummies*	
	β^{SIGN}	$\beta^{NLLS\,***}$	β^{SIGN}	$\beta^{NLLS\,***}$
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.

*** Columns 2 and 4 are taken from Barro and Sala-i Martin (1991).

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 nonrobustness 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 A.1], 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 introduce inference tools that can be associated with the Monte Carlo based system presented in Coudin and Dufour (2007): 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 the degree of confidence one may have in that particular value. It extends the confidence distribution concept to multidimensional parameters and relies on a reinterpretation of the Fisher fiducial distributions. The parameter values the less 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 present general properties of sign-based estimators (invariance, median unbiasedness) and the conditions under which consistency and asymptotic normality hold. In particular, we show 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 arbitrary 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. We present two illustrative applications of such cases. In the first one, we estimate a drift parameter on the Standard and Poor's Composite Price Index, using the 1928-1987 period and various shorter subperiods. In the second one, we provide robust estimates for the β-convergence between the levels of per capita personal income across U.S. States occurred between 1880 and 1988.

Appendix

A. Proofs

Proof of Proposition 4.1. We show that the sets $M1$ and $M2$ are equal with probability one. First, we show that if $\hat{\beta} \in M2$ then it belongs to $M1$. Second, we show that if $\hat{\beta}$ does not belong to $M2$, neither it belongs to $M1$.

If $\hat{\beta} \in M2$ then,

$$D_S(\hat{\beta}) \leq D_S(\beta), \ \forall \beta \in \mathbb{R}^p, \tag{A.1}$$

hence

$$P_\beta[D_S(\hat{\beta}) \leq D_S(\beta)] = 1, \quad \forall \beta \in \mathbb{R}^p \tag{A.2}$$

and $\hat{\beta}$ maximizes the p-value. Conversely, if $\hat{\beta}$ does not belong to $M1$, 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 $M2$. □

Proof of Proposition 5.2. Consider $\hat{\beta}(y, X, u)$ the solution of problem (4.4) which is assumed to be unique, let β_0 be the true value of the parameter β and suppose that $u \sim -u$. Equation (5.12) 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 (5.13) implies that

$$\begin{aligned} \hat{\beta}(y, X, u) &= \hat{\beta}(y - X\beta_0, X, u) + \beta_0 \\ &= \hat{\beta}(u, X, u) + \beta_0. \end{aligned} \tag{A.6}$$

Finally, (A.5) and (A.6) entail $\mathrm{Med}(\hat{\beta}(y, X, u) - \beta_0) = 0$. □

Proof of Theorem 6.1. We consider the stochastic process $W = \{W_t = (y_t,\ x_t') : \Omega \to \mathbb{R}^{p+1}\}_{t=1,2,...}$ defined on the probability space $(\Omega, \mathcal{F}, P)$. We denote

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

The proof of consistency follows four classical steps. First, $\frac{1}{n}\sum_t q_t(W_t, \beta) - E[q_t(W_t, \beta)]$ 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 $\{q_{tk}(W_t, \beta),\ k = 1, \ldots, p\}_{t=1,2,...}$. 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 $E|q_{tk}(W_t, \beta)|^{r+\delta} < \infty$ 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 $\{q_{tk}(W_t, \beta)\}_{t=1,2,...}$. It follows $\forall \beta \in \Theta$,

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

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

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

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 $(\Omega, \mathcal{P}, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B})$, $\forall t,\ \beta \in \Theta,\ \rho$, where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}$ and finally, that $\sup_t \ Eq_{tk}(W_t)^{\xi} < \infty$ 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 $P_t[u_t = x_t\beta] = 0,\ \forall \beta$, uniformly in t. Then, q_{tk} is continuous in β everywhere except on a P_t-negligeable set. Finally, since q_{tk} is L_1-bounded and uniformly integrable, condition A6 holds.

The generic law of large numbers (GWLLN) implies:

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

(b) $\quad \displaystyle\sup_{\beta \in \Theta} \left| \frac{1}{n} \sum_{t=0}^{n} q_t(W_t,\ \beta) - Eq_t(W_t,\ \beta) \right| \to 0$

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

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

Let us define

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

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

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 $\epsilon > 0$, $\delta > 0$ and $n \geq N_0$, we have:

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

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

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

hence,

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

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

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

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

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

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

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

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 2.1, 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 $\Omega_n(\beta)$ defines a norm for each β (condition 6.6), this assertion is equivalent to

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

and
$$\lim_{n\to\infty}\left|E\left[\frac{1}{n}\sum_t s(u_t-x_t'\delta)x_t'\delta\right]\right|=0\Rightarrow\delta=0,\ \delta\in\mathbb{R}^p. \tag{A.14}$$
Let $A(\delta)=E[\frac{1}{n}\sum_t s(u_t-x_t'\delta)x_t|x_1,\ldots,x_n]$. Then,
$$E[A(\delta)]=E\left[\frac{1}{n}\sum_t s(u_t-x_t'\delta)x_t\right]=E\left\{E\left[\frac{1}{n}\sum_t s(u_t-x_t'\delta)x_t|x_1,\ldots,x_n\right]\right\}.$$
Note that
$$E[s(u_t-x_t'\delta)|x_1,\ldots,x_n]=2\left[\frac{1}{2}-\int_{-\infty}^{x_t'\delta}f_t(u|x_1,\ldots,x_n)du\right]=-2\int_0^{x_t'\delta}f_t(u|x_1,\ldots,x_n)du]$$
Hence $A(\delta)$ can be developed for $\tau>0$ as
$$\begin{aligned}A(\delta) \;=\; & \frac{2}{n}\sum x_t'\delta\left\{I_{\{|x_t'\delta|>\tau\}}\left[I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta}-f_t(u|x_1,\ldots,x_n)du\right.\right.\\ & \left.+I_{\{x_t'\delta\le 0\}}\int_{x_t'\delta}^{0}f_t(u|x_1,\ldots,x_n)du\right]\\ & +I_{\{|x_t'\delta|\le\tau\}}\left[I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta}-f_t(u|x_1,\ldots,x_n)du\right.\\ & \left.\left.+I_{\{x_t'\delta\le 0\}}\int_{x_t'\delta}^{0}f_t(u|x_1,\ldots,x_n)du\right]\right\}.\end{aligned}$$
Then,
$$\begin{aligned}E[A(\delta)] = \;& E\left\{\frac{2}{n}\sum x_t'\delta\left[I_{\{|x_t'\delta|>\tau\}}\left(I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta}-f_t(u|x_1,\ldots,x_n)du\right.\right.\right.\\ & \left.+I_{\{x_t'\delta\le 0\}}\int_{x_t'\delta}^{0}f_t(u|x_1,\ldots,x_n)du\right)\\ & +I_{\{|x_t'\delta|\le\tau\}}(I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta}-f_t(u|x_1,\ldots,x_n)du\\ & \left.\left.+\ I_{\{x_t'\delta\le 0\}}\int_{x_t'\delta}^{0}f_t(u|x_1,\ldots,x_n)du)\right]\right\}.\end{aligned}$$
Remark that each term in this sum is negative. Hence, $s(E[A(\delta)])\le 0$ and $|E[A(\delta)]|=-E[A(\delta)]$, and
$$\begin{aligned}|E(A)| \;=\; & E\left[\frac{2}{n}\sum x_t'\delta I_{\{|x_t'\delta|>\tau\}}\left(I_{\{x_i'\delta>0\}}\int_0^{x_t'\delta}f_t(u|x_1,\ldots,x_n)du\right.\right.\\ & \left.\left.-I_{\{x_t'\delta\le 0\}}\int_{x_t'\delta}^{0}f_t(u|x_1,\ldots,x_n)du\right)\right]\\ & +E\left[\frac{2}{n}\sum x_t'\delta I_{\{|x_t'\delta|\le\tau\}}\left(I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta}f_t(u|x_1,\ldots,x_n)du\right.\right.\end{aligned}$$

$$
\begin{aligned}
& -I_{\{x_t'\delta\leq 0\}}\int_{x_t'\delta}^{0} f_t(u|x_1,\ldots,x_n)du\Bigg)\Bigg] \\
\geq\ & E\left[\frac{2}{n}\sum I_{\{|x_t'\delta|>\tau\}}\left(x_t'\delta I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta} f_t(u|x_1,\ldots,x_n)du\right.\right. \\
& \left.\left. -x_t'\delta I_{\{x_t'\delta\leq 0\}}\int_{x_t'\delta}^{0} f_t(u|x_1,\ldots,x_n)du\right)\right] \qquad \text{(A.15)} \\
\geq\ & E\left\{\frac{2}{n}\sum I_{\{|x_t'\delta|>\tau\}}\left[x_t'\delta I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta} f_t(u|x_1,\ldots,x_n)du\right.\right. \\
-\ & \left.\left. x_t'\delta I_{\{x_t'\delta\leq 0\}}\int_{x_t'\delta}^{0} f_t(u|x_1,\ldots,x_n)du\right][f_t(0|x_1,\ldots,x_n)>f_L]p_1\right\} \qquad \text{(A.16)} \\
\geq\ & p_1 E\left\{\frac{2}{n}\sum I_{\{|x_t'\delta|>\tau\}}\tau f_L d|f_t(0|x_1,\ldots,x_n)>f_L\right\}, \qquad \text{(A.17)} \\
\geq\ & \tau p_1 f_L d\frac{2}{n}\sum P[|x_t'\delta|>\tau|f_t(0|x_1,\ldots,x_n)>f_L)]. \qquad \text{(A.18)}
\end{aligned}
$$

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,

$$\lim_{n\to\infty}|E[A(\delta)]|\geq 2\tau p_1 f_L d\times\liminf_{n\to\infty}P[|x_i'\delta|>\tau|f_i(0|x_1,\ldots,x_n)>f_L]>0,\quad\forall\delta>0, \qquad \text{(A.19)}$$

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 (6.17) is the derivative of $\lim_{n\to\infty}\frac{1}{n}\sum_t E\big[s\big(u_t-x_t'(\beta-\beta_0)\big)x_t\big]$. Then, we check the conditions for applying Theorem 7.2 of Newey-McFadden.

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

$$\frac{1}{n}\sum_{t=0}^{n} E\big[s\big(u_t - x_t'(\beta-\beta_0)\big)x_t\big] \tag{A.20}$$

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

$$\frac{\partial}{\partial\beta}E\big[x_t s\big(u_t - x_t'(\beta-\beta_0)\big)\big] = E\big[x_t x_t' f_t\big(x_t'(\beta-\beta_0)|x_1,\ldots,x_n\big)\big]. \tag{A.21}$$

And, these conditions imply that

$$L_n(\beta) = \frac{1}{n}\sum_{t=1}^{n} E\big[x_t x_t' f_t\big(x_t'(\beta-\beta_0)|x_1,\ldots,x_n\big)\big] \tag{A.22}$$

converges uniformly in β to $L(\beta)$. Uniform convergence entails that $\lim_n \frac{1}{n}\sum_{t=0}^{n} E\big[s\big(u_t - x_t'(\beta-\beta_0)\big)x_t\big]$ 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 (2.1, 6.1-6.6), we have to check that the following conditions hold:

(i) zero is attained at the limit by β_0;

(ii) the limiting objective function is differentiable at β_0 with derivative $L(\beta_0)$ such that $L(\beta_0)\Omega L(\beta_0)'$ is nonsingular;

(iii) β_0 is an interior point of Θ;

(iv) $\sqrt{n}Q_n(\beta_0) \to \mathcal{N}(0, J)$

(v) for any $\delta_n \to 0,\ \sup_{||\beta-\beta_0||} \sqrt{n}||Q_n(\beta) - Q_n(\beta_0) - EQ(\beta)||/(1+\sqrt{n}||\beta-\beta_0||) \xrightarrow{p} 0$

Condition (i) is fulfilled by the moment condition 2.1. 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 $\{u_t, X_t\}_{t=1,2,\ldots}$ and conditions 2.1, 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):

$$\sqrt{n}J_n^{-1/2}Q_n(\beta_0) \to N(0, I_p) \tag{A.23}$$

where $J_n = \operatorname{var}\left[\frac{1}{\sqrt{n}}\sum_1^n s(u_i)x_i\right]$. 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

$$\sqrt{n}S_n^{-1/2}\big(\hat{\beta}(\Omega_n) - \beta_0\big) \rightarrow \mathcal{N}(0, I_p).$$

The asymptotic covariance matrix S is given by the limit of

$$S_n = [L_n(\beta_0)\Omega_n(\beta_0)L_n(\beta_0)]^{-1}L_n(\beta_0)\Omega_n(\beta_0)J_n\Omega_n(\beta_0)L_n(\beta_0)[L_n(\beta_0)\Omega_n(\beta_0)L_n(\beta_0)]^{-1}.$$

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

$$\sqrt{n}S_n^{-1/2}\big(\hat{\beta}(\hat{J}_n^{-1}) - \beta_0\big) \rightarrow \mathcal{N}(0, I_p)$$

with

$$S_n = [L_n(\beta_0)\hat{J}_n^{-1}L_n(\beta_0)]^{-1}.$$

When the mediangale Assumption (2.2) 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

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

□

A.1. Detailed empirical results: concentrated statistic and projected p-value graphics

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.

References

ABDELKHALEK, T., AND J.-M. DUFOUR (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.

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 X. SALA-I MARTIN (1991): “Convergence Across States and Regions,” *Brookings Papers on Economic Activity*, 1991(1), 107–182.

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

BUCHINSKY, M. (1995): “Estimating the Asymptotic Covariance Matrix for Quantile Regression Models,” *Journal of Econometrics*, 68(2), 303–338.

——— (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 *Encyclopedia of Statistical Sciences*, ed. by S. Kotz, and N. L. Johnson, vol. 3. Wiley.

CHERNOZHUKOV, V., C. HANSEN, AND M. JANSSON (2008): “Finite Sample Inference for Quantile Regression Models,” *Journal of Econometrics*, forthcoming.

CHERNOZHUKOV, V., E. TAMER, AND H. HONG (2006): “Inference on Parameter Sets in Econometric Models,” Discussion paper 06-18, M.I.T. Department of Economics.

COUDIN, E., AND J.-M. DUFOUR (2007): “Finite-Sample Distribution-Free Inference in Linear Median Regressions under Heteroskedasticity and Nonlinear Dependence of Unknown Form,” Discussion paper, McGill University, CIRANO and CIREQ, and CREST (n°2007-38).

DUFOUR, J.-M. (1990): “Exact Tests and Confidence Sets in Linear Regressions with Autocorrelated Errors,” *Econometrica*, 58(2), 475–494.

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

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

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

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

DUFOUR, J.-M., AND M. TAAMOUTI (2005): “Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments,” *Econometrica*, 73(4), 1351–1365.

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

EFRON, B. (1998): “R. A. Fisher in the 21st Century,” *Statistical Science*, 13(2), 95–122.

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., D. HSIEH, AND G. TAUCHEN (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," Discussion paper, Department of Statistics, Colorado State University.

HANSEN, L. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50(4), 1029–1054.

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

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

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

HOROWITZ, J. (1998): "Bootstrap Methods for Median Regression Models," *Econometrica*, 66(6), 1327–1352.

JAECKEL, L. (1972): "Estimating Regression Coefficients by Minimizing the Dispersion of the Residuals," *The Annals of Mathematical Statistics*, 40(5), 1449–1458.

JOHNSON, N., S. KOTZ, AND C. READ (eds.) (1983): *Encyclopedia of Statistical Sciences*, vol. 3. John Wiley & Sons, New York.

JURECKOVA, J. (1971): "NonParametric Estimate of Regression Coefficients.," *The Annals of Mathematical Statistics*, 42(4), 1328–1338.

KOENKER, R., AND G. BASSETT, JR. (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.

LINTON, O., AND Y.-J. WHANG (2007): "The quantilogram: with an application to evaluating directional predictability," *Journal of Econometrics*, 141(1), 250–282.

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

NEYMAN, J. (1941): “Fiducial argument and the theory of confidence intervals,” *Biometrika*, 32(2), 128–150.

PARZEN, M., L. WEI, AND Z. YING (1994): “A Resampling Method Based on Pivotal Estimating Functions,” *Biometrika*, 81, 341–350.

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

——— (1984): “Least Absolute Deviations Estimation for the Censored Regression Model,” *Journal of Econometrics*, 25(3), 303–325.

SCHWEDER, T., AND N. L. HJORT (2002): “Confidence and Likelihood,” *Scandinavian Journal of Statistics*, 29(2), 309–332.

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

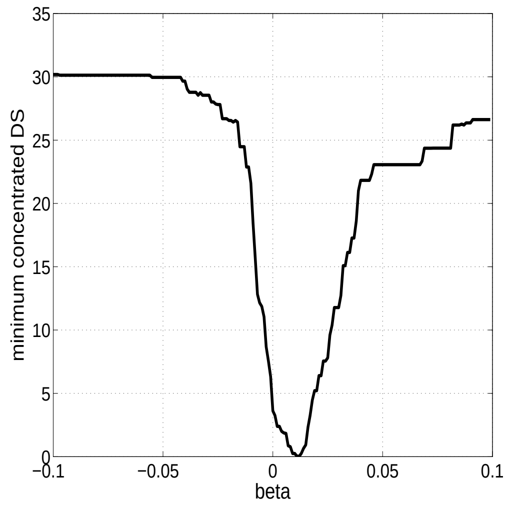
SUNG, N. K., G. STANGENHAUS, AND H. DAVID (1990): “A Cramér-Rao Analogue for Median-Unbiased Estimators,” *Trabajos de Estadistica*, 5, 83–94.

VALÉRY, P., AND J.-M. DUFOUR (2004): “Finite-Sample and Simulation-Based Tests for Stochastic Volatility Model,” Discussion paper, Université de Montréal, CIRANO, CIREQ.

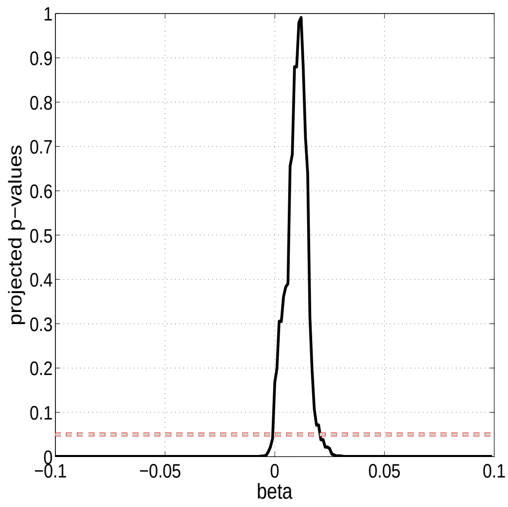
WEISS, A. (1991): “Estimating Nonlinear Dynamic Models Using Least Absolute Error Estimation,” *Econometric Theory*, 7(1), 46–68.

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

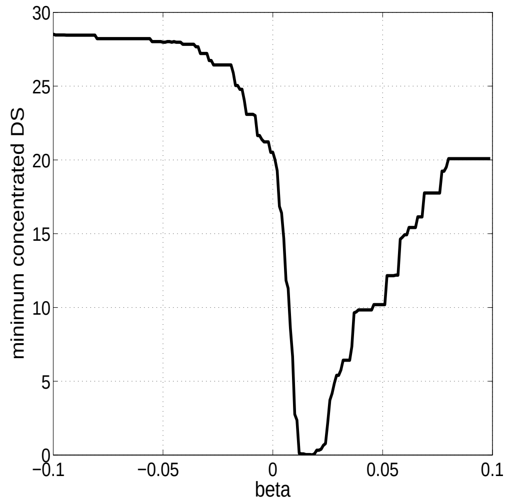
ZHAO, Q. (2001): “Asymptotically Efficient Median Regression in Presence of Heteroskedasticity of Unknown Form,” *Econometric Theory*, 17, 765–784.



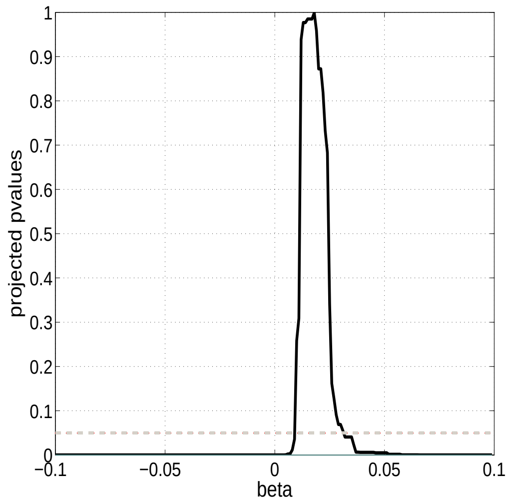
(a) Basic equation: 1880-1900: concentrated DS



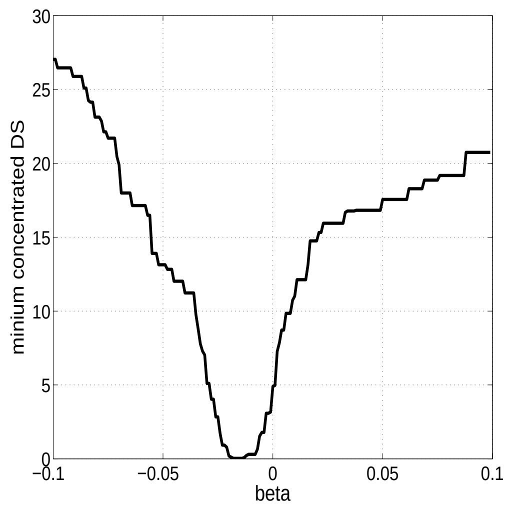
(b) Basic equation: 1880-1900: projected p-value



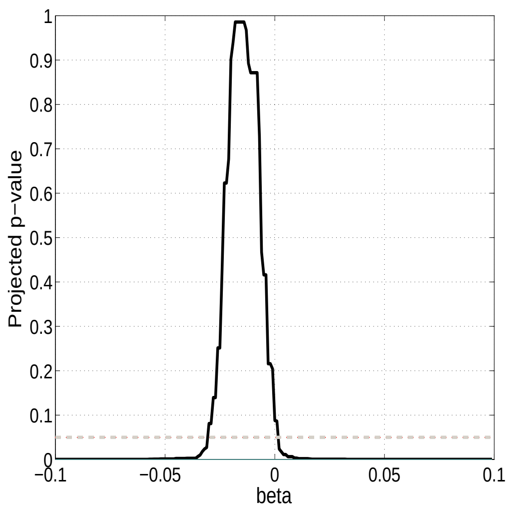
(c) Basic equation: 1900-20: concentrated DS



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

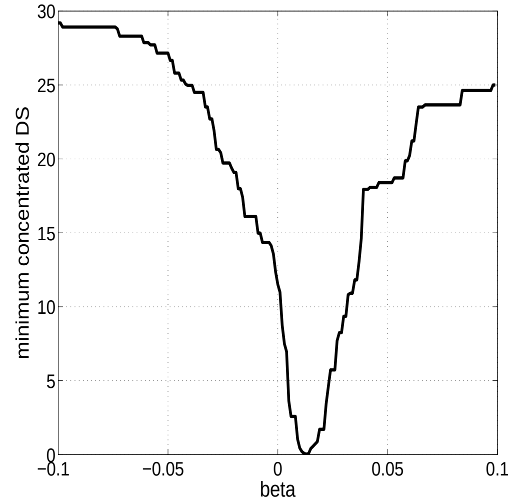


(e) Basic equation: 1920-30: concentrated DS

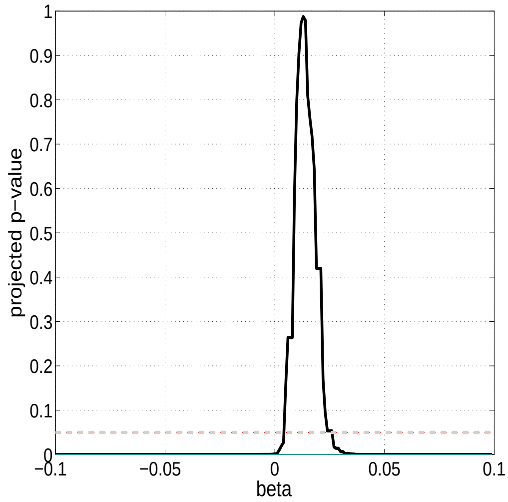


(f) Basic equation: 1920-30: projected p-value

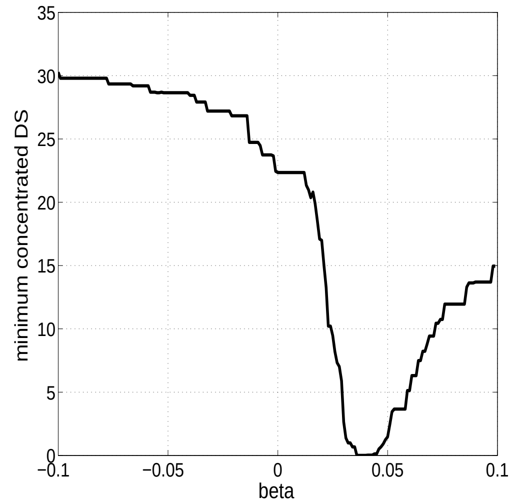
Figure 5. Concentrated statistics and projected p-values (1880-1930)



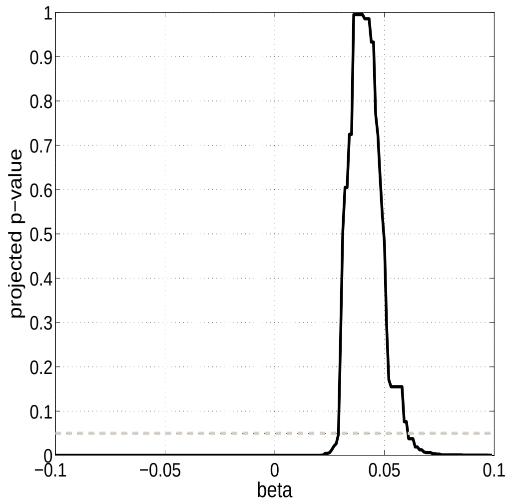
(a) Basic equation: 1930-40: concentrated DS



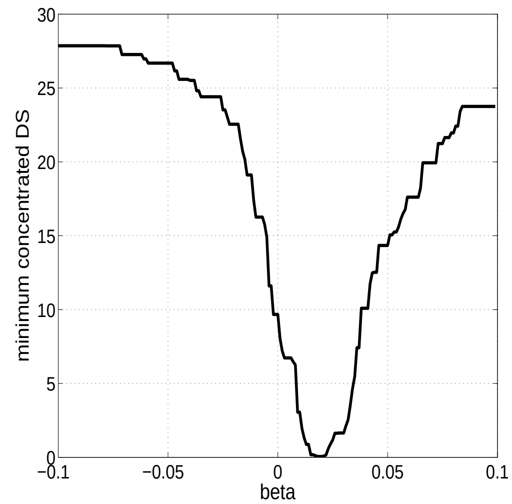
(b) Basic equation: 1930-40: projected p-value



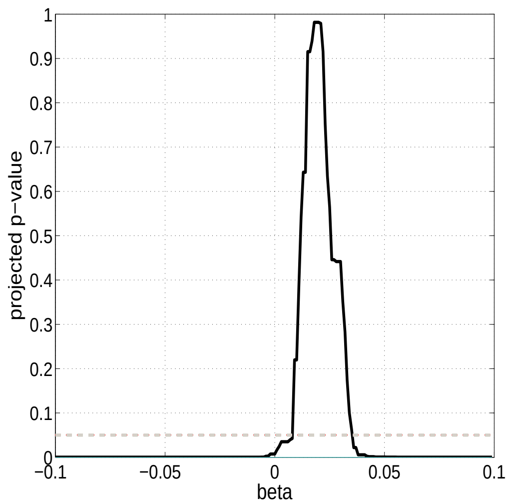
(c) Basic equation: 1940-50: concentrated DS



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

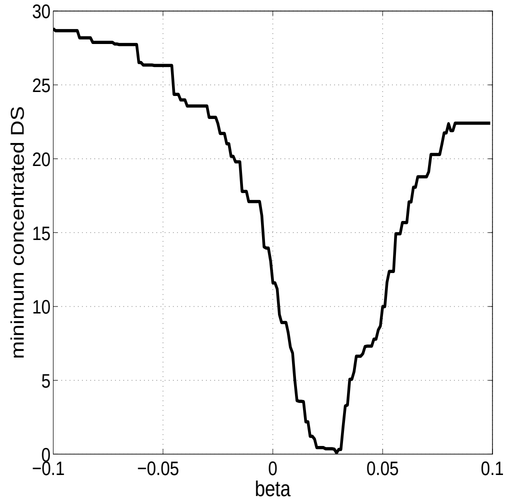


(e) Basic equation: 1950-60: concentrated DS

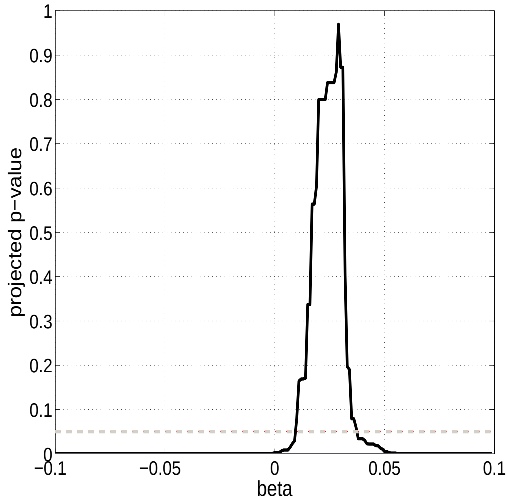


(f) Basic equation: 1950-60: projected p-value

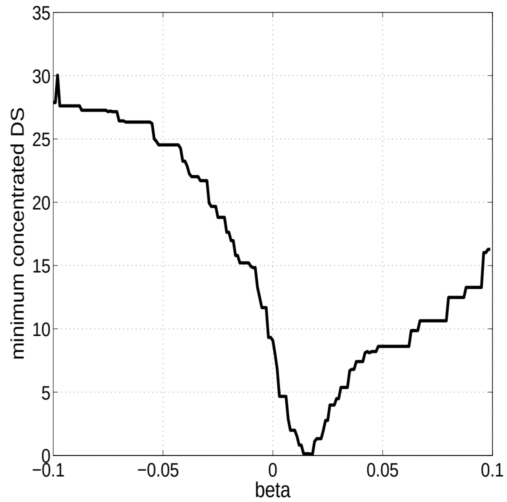
Figure 6. Concentrated statistics and projected p-values (1880-1930)



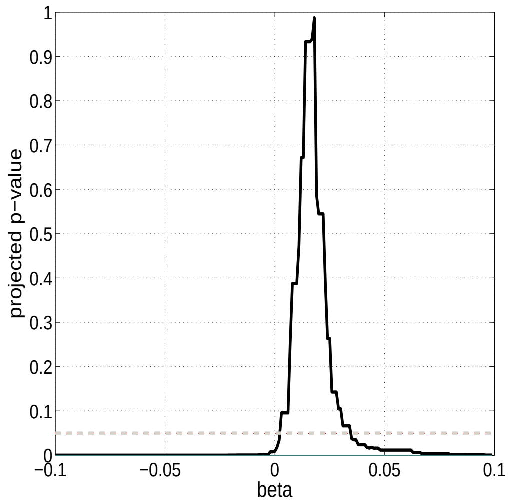
(a) Basic equation: 1960-70: concentrated DS



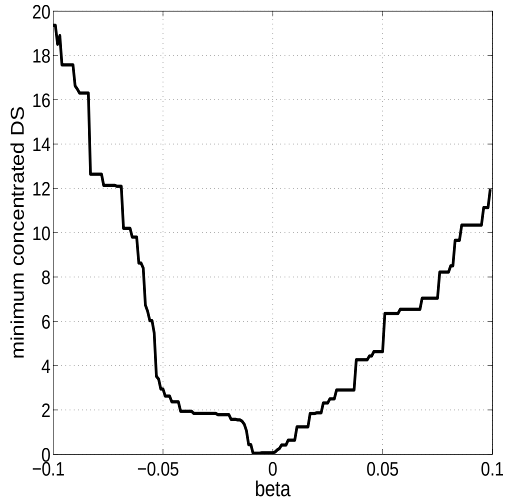
(b) Basic equation: 1960-70: projected p-value



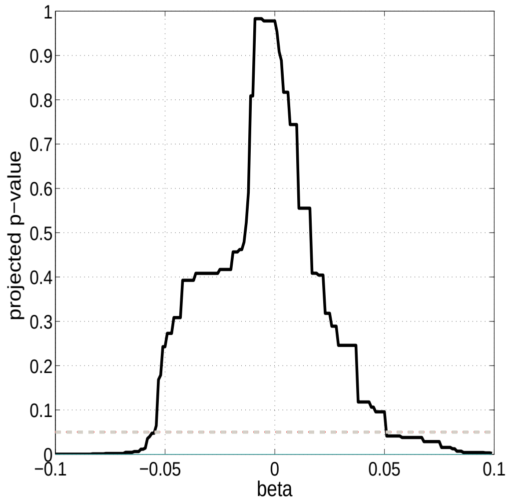
(c) Basic equation: 1970-80: concentrated DS



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



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



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

Figure 7. Concentrated statistics and projected p-values (1880-1930)