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

Elise Coudin † Jean-Marie Dufour ‡
INSEE-CREST McGill University

November 2019

This paper is forthcoming in *Econometric Reviews*.

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

† Centre de recherche en économie et statistique (CREST-ENSAE) and INSEE, Paris, France. E-mail: elise.coudin@ensae.fr

‡ William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 414, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 6071; FAX: (1) 514 398 4800; e-mail: `jean-marie.dufour@mcgill.ca`. Web page: `http://www.jeanmariedufour.com`

ABSTRACT

We study the problem of estimating the parameters of a linear median regression without any assumption on the shape of the error distribution – including no condition on the existence of moments – allowing for heterogeneity (or heteroskedasticity) of unknown form, noncontinuous distributions, and very general serial dependence (linear and nonlinear). This is done through a *reverse inference approach*, based on a distribution-free sign-based testing theory, from which confidence sets and point estimators are subsequently generated. We propose point estimators, which have a natural association with confidence distributions. These estimators are based on maximizing test p-values and inherit robustness properties from the generating distribution-free tests. Both finite-sample and large-sample properties of the proposed estimators are established under weak regularity conditions. We show that they are median unbiased (under symmetry and estimator unicity) and possess equivariance properties. Consistency and asymptotic normality are established without any moment existence assumption on the errors. A Monte Carlo study of bias and RMSE shows sign-based estimators perform better than LAD-type estimators in various heteroskedastic settings. We illustrate the use of sign-based estimators on an example of β-convergence of output levels across U.S. States.

Key words: sign-based methods; median regression; test inversion; Hodges-Lehmann-type estimators; p-value function; least absolute deviation estimators; quantile regressions; sign test; simultaneous inference; Monte Carlo tests; projection methods; non-normality; heteroskedasticity; serial dependence; GARCH; stochastic volatility.

Journal of Economic Literature classification: C13, C12, C14, C15.

1. Introduction

The Laplace-Boscovich median regression is an attractive approach because it can yield estimators [such as the least absolution deviation (LAD) estimator] and tests which are considerably more robust to non-normality and outliers than least-squares methods; see Dodge (1997).[1] An important reason why such methods yield more robust inference comes from the fact that hypotheses about moments are not generally testable in nonparametric setups, while hypotheses about quantiles remain testable under similar conditions [see Bahadur and Savage (1956), Dufour (2003), Dufour, Jouneau and Torrès (2008)]. In such models, Coudin and Dufour (2009) developed provably valid tests, under weak conditions, which allow for very general forms of heterogeneity (or heteroskedasticity). Hypothesis testing is indeed intrisically "simpler" than point estimation, because the null hypothesis typically sets a number of model coefficients, which in turn can avoid various distributional complications. By contrast, designing a "consistent" point estimate requires global assumptions which may not be met at specific points of the parameter space. This type of situation is similar to what happens in models where identification may fail: in such cases, hypothesis testing problems can be immune to identification failure, for example because unidentifiable parameters are fixed by the null hypothesis, so it is better to start with hypothesis testing [see Dufour (1997, 2003)]. Indeed, in the weak identification literature, identification-robust tests (such as Anderson-Rubin tests) are derived first and then "inverted" to build identification-robust confidence sets (though typically not point estimates).

As originally suggested by Hodges and Lehmann (1963) for the problem of estimating a location parameter, it is also possible to derive point estimates from distribution-free tests (*e.g.*, rank tests) by finding the parameter value which is "least rejectable". This method stands in sharp contrast with the common approach which consists in first finding a point estimator (*e.g.*, some consistent estimator), derive a distributional theory for the estimator, and then build tests and confidence sets based on the latter distributional theory. The usual process is in fact "reversed": we start from hypothesis tests (based on weak nonparametric distribution assumptions), build confidence sets, and then derive point estimates. We call this process the *reverse inference approach*.

In this paper, we propose to use such an approach to derive robust point-estimation methods. Specifically, we study the problem of estimating the parameters of a linear median regression without any assumption on the shape of the error distribution – allowing for heterogeneity (or heteroskedasticity) of unknown form, noncontinuous distributions, and very general serial dependence (linear and nonlinear). This is done through a reverse inference approach, which starts from a distribution-free testing theory [Coudin and Dufour (2009)], subsequently exploited to derive point estimators. Using the tests proposed in Coudin and Dufour (2009), the estimation problem is tackled in two complementary ways.

First, we show how *confidence distributions* for model parameters [Schweder and Hjort (2002), Xie and Singh (2013)] can be applied in such a context. Such distributions – which can be interpreted as a form of *fiducial inference* [Fisher (1930), Buehler (1983), Efron (1998), Hannig (2006)] – provide a frequency-based method for associating probabilities with subsets of the parameter space (like posterior distributions do in a Bayesian setup) without the introduction of a prior distribution. In the one-dimensional model, the confidence distribution is defined as a distribution whose quantiles span all the possible confidence intervals [Schweder and Hjort (2002)]. In this paper, we consider *generalized confidence distributions* applicable to multidimensional parameters, and we suggest the use of a projection technique for confidence inference on individual model parameters. The latter are exact – in the sense that the parameters considered are covered with known probabilities (or larger) –

[1] This holds also for quantile regressions [Koenker and Bassett (1978), Koenker (2005)], which can be viewed as extensions of median regression.

under the mediangale assumption considered in Coudin and Dufour (2009). Further, if more general linear dependence is allowed, the proposed method remains asymptotically valid.

Second, we propose point estimates, which bear a natural association with the above confidence distributions. These Hodges-Lehmann-type estimators are based on maximizing test p-values and inherit several robustness properties from the distribution-free tests used to generate them. Both finite-sample and large-sample properties are established under weak regularity conditions. We show they are median unbiased (under symmetry and estimator unicity) and possess equivariance properties with respect to linear transformations of model variables. Consistency and asymptotic normality are established without any moment existence assumption on the errors, allowing noncontinuous distributions, heterogeneity and general serial dependence of unknown form. These conditions are weaker or similar to conditions used in the literature on LAD estimators; see Bassett and Koenker (1978), Bloomfield and Steiger (1983), Powell (1984), Knight (1989, 1998), Phillips (1991), Pollard (1991), Portnoy (1991), Weiss (1991), Fitzenberger (1997), El Bantli and Hallin (1999), Zhao (2001), Koenker (2005), and the references therein.

In practice, for any given sample size, the sign transform enables one to construct test statistics with known nuisance-parameter-free distributions without additional parametric restrictions. Realized p-value functions are then constructed by testing hypotheses of the form $H_0(\beta_0) : \beta = \beta_0,$ where β is the vector of the regression coefficients with Monte Carlo test methods and projection techniques described in this context in Coudin and Dufour (2009). [2] For each component, a projected p-value function provides a representation of the evidence for each possible value of this component. Using the above p-values (as a function of β_0), we then derive estimators by taking a value β_0 which is "least rejected" by the test procedure. First applied to the Wilcoxon's signed rank-statistic for estimating a shift or a location, this principle was adapted to regression models by deriving so-called *R*-estimators from rank or signed-rank statistics [Jureckova (1971), Jaeckel (1972), Koul (1971)]. We will see that these estimators can alternatively be computed by minimizing quadratic forms of the constrained signs. The class of sign-based estimators we propose includes as special cases the *sign estimators* derived by Boldin, Simonova and Tyurin (1997) from locally most powerful sign tests in linear regressions with *i.i.d.* errors and fixed regressors. A major advantage of signs over ranks consists in dealing transparently with heteroskedastic (or heterogeneous) disturbances. Many heteroskedastic and possibly dependent schemes are covered and, in presence of linear dependence, a HAC-type correction for heteroskedasticity and autocorrelation can be included in the criterion function. The conjunction of sign-based tests, projection-based confidence regions, and sign-based estimators thus provides a complete system of inference.

We study the performance of the proposed estimators in a Monte Carlo study which allows for several non-Gaussian and heteroskedastic setups. We find that sign-based estimators are competitive (in terms of bias and RMSE) when errors are *i.i.d.*, while they are substantially more reliable than usual methods (LS, LAD) when heterogeneity or serial dependence is present in the error term. Finally, we illustrate their use in practice on a exercise which revisits β-convergence of output levels across U.S. States.

The paper is organized as follows. Section 2 recalls the framework and the main results of Coudin and Dufour (2009) used in the present paper. Section 3 is dedicated to confidence distributions and p-value functions. In section 4, we define the proposed family of sign-based estimators. The finite-sample properties of the sign-based estimators are studied in section 5, while their asymptotic properties are considered in section 6. In section 7, we present the results of our simulation study of

[2] See also Dwass (1957), Barnard (1963), Dufour (1990, 1997), Dufour and Kiviet (1998), Abdelkhalek and Dufour (1998), Dufour and Jasiak (2001), Dufour and Taamouti (2005). And for an alternative finite-sample inference exploiting a quantile version of the same sign pivotality result, which holds if the observations are X-conditionally independent, see Chernozhukov, Hansen and Jansson (2009).

bias and RMSE. The empirical application is reported in section 8 and conclusion in section 9. The proofs and some additional numerical results are available in a separate *Technical Appendix*.

2. Framework

We use the same framework as Coudin and Dufour (2009) which we briefly summarize below. We consider a stochastic process $\{(y_t, x_t') : \Omega \to \mathbb{R}^{p+1} : t = 1, 2, \ldots\}$ defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, 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 $y = (y_1, \ldots, y_n)' \in \mathbb{R}^n$ will denote the dependent variable vector, $X = [x_1, \ldots, x_n]' \in \mathbb{R}^{n\times p}$ the $n \times p$ matrix of explanatory variables, which may be random or fixed, and $u = (u_1, \ldots, u_n)' \in \mathbb{R}^n$ the disturbance vector. We will consider assumptions on the signs $s(u_1), \ldots, s(u_n)$ of model errors, where the sign function $s : \mathbb{R} \to \{-1, 0, 1\}$ 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 $a \notin A$. $s(u) = [s(u_1), \ldots, s(u_n)]'$ denotes the vector of the signs of the components.

We will consider as sign-based statistics general quadratic forms involving the vector $s(y - X\beta_0)$ of the constrained signs (*i.e.*, the signs aligned with respect to $X\beta_0$):

$$D_S[\beta_0, \bar{\Omega}_n(\beta_0)] = s(y - X\beta_0)'X\Omega_n\big[s(y - X\beta_0), X\big]X's(y - X\beta_0) \tag{2.2}$$

where $\bar{\Omega}_n(\beta_0) = \Omega_n\big[s(y - X\beta_0), X\big]$ is a $p \times p$ positive definite weight matrix which may depend on the constrained signs.

Coudin and Dufour (2009) derive a finite-sample distribution-free inference system for testing $H_0(\beta_0) : \beta = \beta_0$ **vs.** $H_1(\beta_0) : \beta \neq \beta_0$ in model (2.1) using sign-based statistics under a mediangale assumption. This assumption ensures that sign-based statistics constitute pivotal functions whose distributions conditional on X can be simulated, and exact Monte-Carlo tests can be constructed [Dufour (2006)]. It can be stated in the context of *adapted sequences* $\mathscr{S}(\mathbf{v}, \mathscr{F}) = \{v_t, \mathscr{F}_t : t = 1, 2, \ldots\}$ where v_t is any measurable function of $W_t = (y_t, x_t')'$, $\mathscr{F}_t$ is a σ-field in Ω, $\mathscr{F}_s \subseteq \mathscr{F}_t$ for $s < t$, $\sigma(W_1, \ldots, W_t) \subset \mathscr{F}_t$ and $\sigma(W_1, \ldots, W_t)$ is the σ-algebra spanned by $W_1, \ldots, W_t$.

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

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

Besides nonnormality (including no condition on the existence of moments), this assumption allows for heterogeneity (or heteroskedasticity) of unknown form, heavy-tailed distributions, noncontinuous distributions, and general forms of nonlinear serial dependence, including GARCH-type and stochastic volatility of unknown order. It does not, however, cover "linear serial dependence" such as an ARMA process on u_t. Hence, our asymptotic results will rely on the following standard moment condition. Note however, that this condition is not required by confidence distributions, which only depend of the distributional theory of sign tests.

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

3. Confidence distributions

In the one-parameter model, statisticians have defined the confidence distribution notion which 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 this 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 this 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 sign-based regressions

Schweder and Hjort (2002) defined the confidence distribution for the real parameter β as a distribution which depends on the observations (y, x), whose cumulative distribution function evaluated at the true value of β is uniform irrespective of 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*

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

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

Here $(-\infty,\ CD^{-1}(\alpha)]$ is a one-sided stochastic confidence interval with coverage probability α.[3] The realized confidence $CD(\beta_0; y, x)$ is the p-value of the one-sided hypothesis $H_0^*(\beta_0):\ \beta \le \beta_0$ versus $H_1^*(\beta_0):\ \beta > \beta_0$ when the observed data are $y,\ x$. The realized p-value when testing $H_0(\beta_0):$ $\beta = \beta_0$ versus $H_1(\beta_0):\ \beta \ne \beta_0$ is $2\min\{CD(\beta_0), 1 - CD(\beta_0)\}$.[4] Hence, tests and confidence intervals on β are contained in the confidence distribution. 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 $S(\beta)$ with cumulative distribution function F which is independent of β and free of any nuisance parameter), $F\big(S(\beta)\big)$ is uniformly distributed and satisfies conditions for providing a confidence distribution. Let $S(\beta)$ be such a continuous real statistic increasing with β with a nuisance-parameter-free distribution. A test of $H_0 : \beta \le \beta_0$ rejects H_0 when $S^{obs}(\beta_0)$ is large, with p-value $\mathbb{P}_{\beta_0}[S(\beta_0) > S^{obs}(\beta_0)]$. Then,

$$\mathbb{P}_{\beta_0}[S(\beta_0) > S^{obs}(\beta_0)] = 1 - F_{\beta_0}[S^{obs}(\beta_0)] = CD(\beta_0) \tag{3.2}$$

[3] For continuous distributions, just note that $P_\beta[\beta \le CD^{-1}(\alpha)] = P_\beta\{CD(\beta) \le CD\big(CD^{-1}(\alpha)\big)\} = P_\beta\{CD(\beta) \le \alpha]\} =$ α. Schweder and Hjort (2002) introduce the notion of "degree of confidence" $CD(\beta_0)$. of the statement $\beta \le \beta_0$ which is equals to the p-value of a test $\beta \le \beta_0$ versus the alternative $\beta > \beta_0$.

[4] Those relations are stated in Lemma 2 of Schweder and Hjort (2002): *the confidence of the statement "$\beta \le \beta_0$" is the degree of confidence $CD(\beta_0)$ for the confidence interval $\big(-\infty, CD^{-1}\big(CD(\beta_0)\big)\big]$, and is equal to the p-value of a test of $H_0^*(\beta_0) : \beta \le \beta_0$ vs. $H_1^*(\beta_0) : \beta > \beta_0$.*

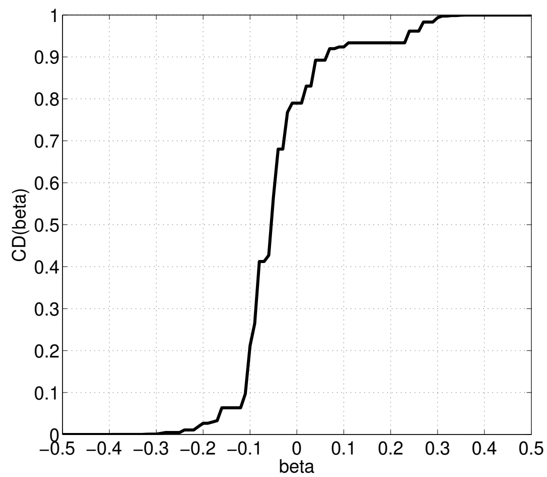


Figure 1. Simulated confidence distribution cumulative function based on SST.

where $F_{\beta_0}(.)$ is the sampling distribution of $S(\beta_0)$ under $\beta = \beta_0$. Consequently, simulated sampling distributions and simulated realized p-values for discrete statistics as proposed by Coudin and Dufour (2009) yield a way to construct simulated confidence distributions.[5]

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

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

$$SST(\beta_0) = \frac{\sum s(y_i - x_i\beta_0)x_i}{(\sum x_i^2)^{1/2}} \tag{3.3}$$

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

$$\widehat{CD}(\beta_0) = 1 - \hat{F}_{\beta_0}[SST(\beta_0)] \tag{3.4}$$

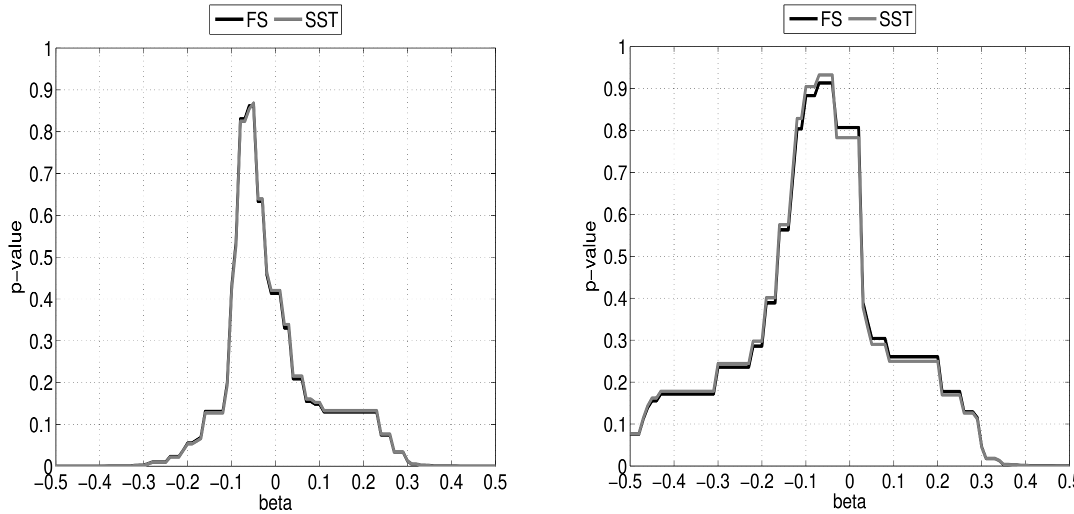
with $\hat{F}_{\beta_0}$ a Monte Carlo estimate of the sampling distribution of SST under $H_0(\beta_0) : \beta = \beta_0$. Figure 1 presents a simulated confidence distribution cumulative function for β, given 200 realizations of (u_i, x_i) based on SST. The Monte Carlo estimate of $\hat{F}_{\beta_0}$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

Realized p-value functions for discrete statistics. Another interesting object is the realized p-value 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

$$\hat{p}_{SST}(\beta_0) = 2\min\{\widehat{CD}_{SST}(\beta_0), 1 - \widehat{CD}_{SST}(\beta_0)\}. \tag{3.5}$$

Consider now the statistic $SF = SST^2$. SF is a pivotal function but not a monotone function of β

[5] Continuous uniform distribution is obtained using a randomization process on ties in Coudin and Dufour (2009).



(a) Example 1: well identified case

(b) Example 2: misspecified case

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

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 simulated p-values based on $D_S[\beta_0,\, \bar{\Omega}_n(\beta_0)]$ as described in Coudin and Dufour (2009). Figures 2 (a) and (b) compare p-value functions based on *SST* and *SF*. Inverting the p-value function allows one to recover half of the confidence distribution and consequently half of the inference results, *i.e.* the two-sided confidence intervals. For example, in Figure 2 (a), $[-.17, .24]$ constitutes a confidence interval with level 90% for both statistics. The p-value function thus provides a summary on the available inference. In particular, it gives a confidence degree to be associated with $\beta = \beta_0$. Finally, the p-value function has an important advantage over the confidence distribution: it is straightforwardly extendable to multidimensional parameters.

The spread of the p-value function is also related to the *model specification* and the *parameter identification*. When the p-value function is flat, one may expect the parameter to be badly identified either because there exists a set of observationally equivalent parameters (p-values are high for a wide set of values), or because there does not exist any value satisfying the model (p-values are small everywhere). To illustrate this point, let us consider another example (example 2) where the first n_1 observations satisfy $y_i = \beta_1 x_i + u_i,\ i = 1, \ldots, n_1,\ (u_i, x_i) \overset{iid}{\sim} \mathrm{N}[0, I_2]$ and the n_2 following ones, $y_i = \beta_2 x_i + u_i,\ i = n_1 + 1, \ldots, n_1 + n_2,\ (u_i, x_i) \overset{iid}{\sim} \mathrm{N}[0, I_2]$, with $\beta_1 = -.5$ and $\beta_2 = .5$. The model $y_i = \beta x_i + u_i,\ i = 1, \ldots, n_1 + n_2$, is misspecified. In Figure 2(b), we notice the spread of the p-value function based on *SF* is large: the set of observationally equivalent β is not reduced to a point.

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

If $p \geq 2$, the confidence distribution notion is not defined anymore. However, simulated realized p-values for testing $H_0(\beta_0)$ can easily be constructed from the *SF* statistic, and more generally from any sign-based statistic which satisfies equation (2.2). Simulated p-values lead to a mapping for which we have a 3-dimensional representation for $p = 2$. Consider the model: $y_i = \beta^1 x_{1i} + \beta^2 x_{2i} + u_i,\ i = 1, \ldots, n,\ (u_i, x_{1i}, x_{2i}) \overset{iid}{\sim} \mathrm{N}[0, I_3]$, $\beta = (\beta^1, \beta^2) = (0, 0)'$, $y = (y_1, \ldots, y_n)'$,

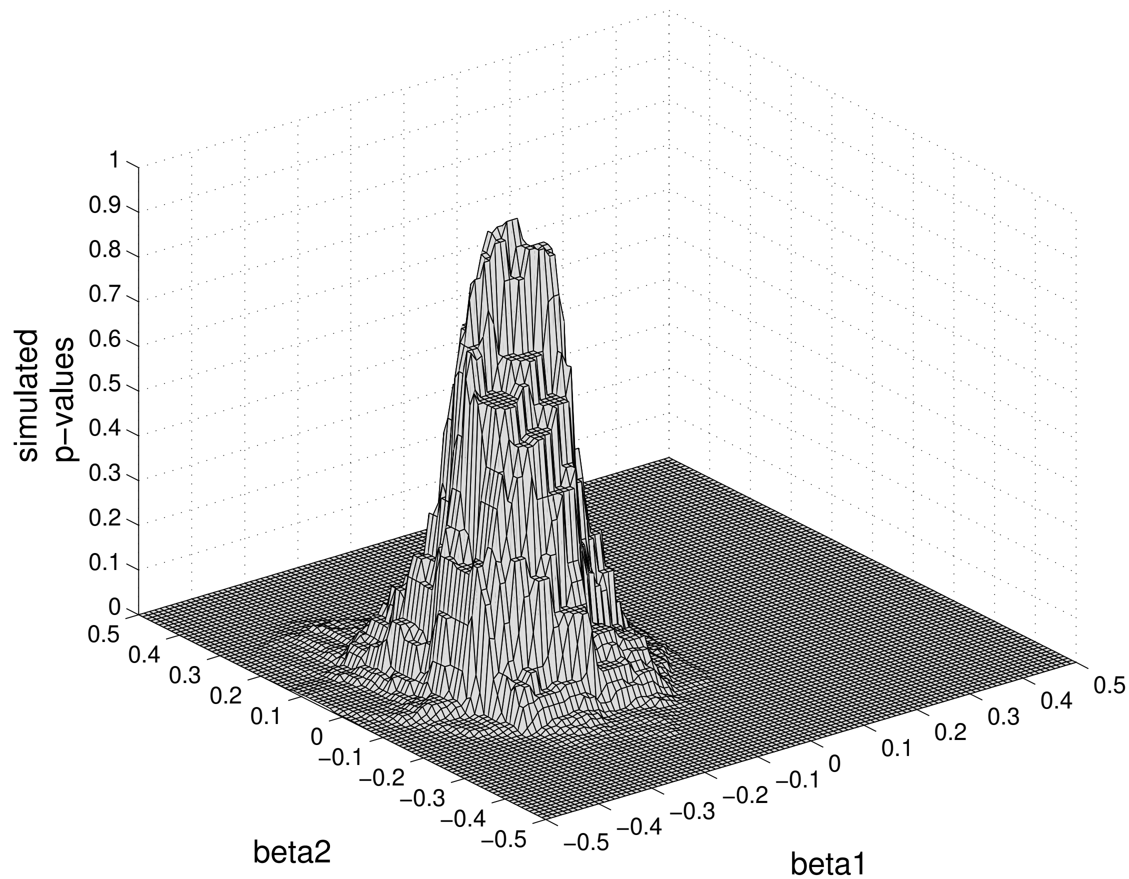


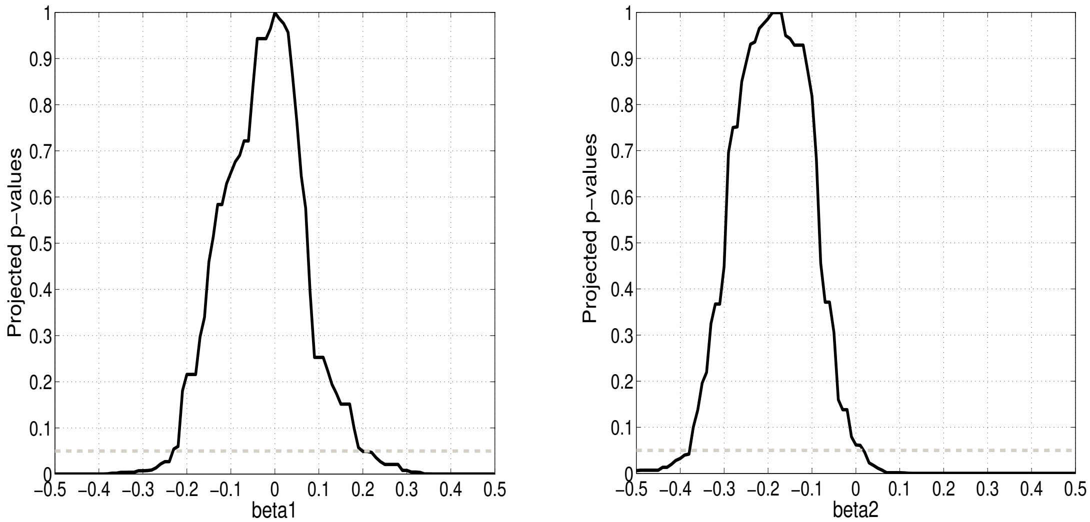
Figure 3. Simulated p-value functions based on SF ($n = 200$, $N = 9999$).

$u = (u_1, \ldots, u_n)'$, $x_1 = (x_{11}, \ldots, x_{1n})'$, $x_2 = (x_{21}, \ldots, x_{2n})'$ and $X = [x_1, x_2]$. Let $D_S[\beta, (X'X)^{-1}] = s(y - X\beta)'X(X'X)^{-1}X's(y - X\beta)$. 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

$$C_{1-\alpha}(\beta) = \{\beta | \tilde{p}_N^{D_S}(\beta_0) \geq \alpha\}, \tag{3.6}$$

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

For higher dimensions ($p > 2$), one can consider projection-based realized p-value functions for each individual component of the parameter, in a way similar way than projection-based confidence sets [Dufour (1990, 1997), Dufour and Kiviet (1998), Abdelkhalek and Dufour (1998), Dufour and Jasiak (2001), Dufour and Taamouti (2005)]. 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



(a) Projection-based $p$values for β^1 (b) Projection-based $p$values for β^2

Figure 4. Projection-based p-values.

component β^1 is given by:

$$\text{Proj.}\tilde{p}_N^{\beta^1}(\beta_0^1) = \max_{\beta_0^2 \in \mathbb{R}} \tilde{p}_N^{D_S}[(\beta_0^1, \beta_0^2)]. \tag{3.7}$$

Figure 4 presents projection-based confidence intervals for the individual parameters of the previous two-dimensional example: $[-.22, .21]$ is a 95% (conservative) confidence interval for β^1, while $[-.38, .02]$ is a 95% (conservative) confidence interval for β^2. The hypothesis $\beta^1 = 0$ is accepted at level 5% with p-value 1.0, and the hypothesis $\beta^2 = 0$ is accepted at 5% with p-value .06.

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

$$\mathbb{P}[\tilde{p}_N^{D_S}(\beta_0) \leq \alpha] = \alpha \quad \forall \alpha \text{ such that } \alpha(N+1) \in \mathbb{N}. \tag{3.8}$$

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:

$$\frac{I(\alpha(N+1)-1)}{N+1} \leq \mathbb{P}[\tilde{p}_N^{D_S}(\beta_0) \leq \alpha] < \alpha \quad \forall \alpha \text{ such that } \alpha(N+1) \notin \mathbb{N}. \tag{3.9}$$

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 ensures 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-value. Estimators obtained by this type of test inversion constitute multidimensional extensions of the Hodges-Lehmann principle.

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

Hodges and Lehmann (1963) presented a general principle to derive estimators by test inversion; see also Johnson, Kotz and Read (1983). Suppose $\mu \in \mathbb{R}$ and $S(\mu_0, W)$ is a statistic for testing $\mu = \mu_0$ against $\mu > \mu_0$ based on the observations W. Suppose further that $S(\mu, W)$ is nondecreasing in the scalar μ. Given a known central value of $S(\mu_0, W)$, say $m(\mu_0)$ [for example $E_W S(\mu_0, W)$], the test rejects $\mu = \mu_0$ whenever the observed S is larger than, say, $m(\mu_0)$. In this 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 $S(\mu^*, W)$ 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 $S(\beta_0)$. A test based on $S(\beta_0)$ rejects $H_0(\beta_0)$ when $S(\beta_0)$ is larger than a certain critical value which 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 $\mathrm{p}(\beta_0) = G[D_S(\beta_0)|\beta_0]$, where $G(x|\beta_0)$ is the survival function of $D_S(\beta_0)$, *i.e.* $G(x|\beta_0) = \mathbb{P}[D_S(\beta_0) > x]$, the set

$$M_1 = \underset{\beta \in \mathbb{R}^p}{\arg\max}\ \mathrm{p}(\beta) \tag{4.1}$$

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

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

When the distribution of $S(\beta_0)$ 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 $S(\beta_0)$. This point is stated in the following proposition. Let us denote $\bar{F}(x|\beta_0)$ the distribution of $S(\beta_0)$ when $\beta = \beta_0$ and assume this distribution is invariant to β (Assumption 4.1).

Assumption 4.1 INVARIANCE OF THE DISTRIBUTION FUNCTION.

$$\bar{F}(x|\beta_0) = \bar{F}(x) \quad \forall x \in \mathbb{R}^+,\ \forall \beta_0 \in \mathbb{R}^p. \tag{4.2}$$

Let us define:

$$M_1 = \underset{\beta \in \mathbb{R}^p}{\mathrm{argmax}}\ \mathrm{p}(\beta), \quad M_2 = \underset{\beta \in \mathbb{R}^p}{\arg\min}\ S(\beta)\,. \tag{4.3}$$

Then, the following proposition holds.

Proposition 4.1 *If Assumption 4.1 holds, then $M_1 = M_2$ with probability one.*

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

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

equals (with probability one) a Hodges-Lehmann-type 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 implies 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 sign-based estimators studied by Boldin et al. (1997) are solutions of

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

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, weighted versions of sign-based estimators can be more efficient, such as the weighted LAD estimator compared to the LAD estimator [see Zhao (2001)]:

$$\hat{\beta}_n^{DX} \in \underset{\beta \in \mathbf{R}^p}{\operatorname{argmin}}\, s(Y - X\beta)' \tilde{X} (\tilde{X}'\tilde{X})^{-1} \tilde{X}' D' s(Y - X\beta) \tag{4.7}$$

where $\tilde{X} = \mathrm{diag}(d_1, \ldots, d_n) X$ and $d_i > 0$, $i = 1, \ldots, n$. Weighted sign-based estimators which involve optimal estimating functions in the sense of Godambe (2001) are solutions of

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

where $X^* = \mathrm{diag}\big(f_1(0|X), \ldots, f_n(0|X)\big) X$ 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

In some interesting special cases, the sign-based estimators proposed in this paper may be interpreted (at least asymptotically) as GMM estimators which exploit the orthogonality condition between the signs and the explanatory variables [see Honore and Hu (2004)].[6] In settings where only the sign-moment Assumption 2.2 is satisfied, the GMM interpretation of sign-based estimators still applies and implies useful extensions.

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

$$\hat{\beta}_n(\hat{J}_n^{-1}) \in \arg \min_{\beta \in \mathbf{R}^p} s(Y - X\beta)' X [\hat{J}_n(s(Y - X\beta), X)]^{-1} X' s(Y - X\beta), \tag{4.9}$$

[6] Due to the use of the nonlinear p-value transformation (along with the associated finite-sample distributional theory), the GMM interpretation does not stricto sensu generally hold, except possibly through an asymptotic equivalence.

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 (4.9) 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,

$$\hat{\beta}_n^{2S}(\hat{J}_n^{-1}) \in \arg\min_{\beta\in\mathbf{R}^p} s(Y-X\beta)'X[\hat{J}_n(s(Y-X\hat{\beta}_n),X)]^{-1}X's(Y-X\beta). \tag{4.10}$$

The 2-step estimator is not a Hodges-Lehmann-type estimator anymore. However, it is still consistent and share some interesting finite-sample properties with classical sign-based estimators.

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 [see Koenker and Bassett (1978)] and are median-unbiased if the disturbance distribution is symmetric and some additional assumptions on the form of the solution are satisfied. The topology of the argmin set of the optimization problem (4.4) does not possess a simple structure. In some cases, it is reduced to a single point like the empirical median of $2p+1$ observations. In other cases, it is a set. More generally, the argmin set is a union of convex sets but it is not *a priori* either convex nor connected.[7] Despite these complications, the following proposition holds.

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

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

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

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

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

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

[7] To see that it is a union of convex sets just remark that the reciprocal image of n fixed signs is convex.

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

$$\mathrm{Med}(\hat{\beta} - \bar{\beta}) = 0 \tag{5.7}$$

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

6. Asymptotic properties

We demonstrate consistency and asymptotic normality of the proposed sign-based estimators when the parameter is identified under weak assumptions. For reviews of the asymptotic distributional theory of LAD estimators, the reader may consult Bassett and Koenker (1978), Knight (1989), Phillips (1991), Pollard (1991), Portnoy (1991), Weiss (1991), Fitzenberger (1997), Knight (1998), El Bantli and Hallin (1999), and Koenker (2005).

6.1. Identification and consistency

The sign-based estimators (4.4) and (4.10) are consistent under the following set of assumptions. In the sequel, we denote by $\bar{\beta}$ the "true value" of β, and by β_0 any hypothesized value.

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

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

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

Assumption 6.4 REGULARITY OF THE DENSITY.

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

$$\mathbb{P}[f_t(0|X) > f_L] > p_1,\ t = 1, \ldots, n,\ a.s. \tag{6.1}$$

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

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

$$\liminf_{n\to\infty} \frac{1}{n} \sum_t \mathbb{P}[|x_t'\delta| > \tau \,|\, f_t(0|x_1, \ldots, x_n) > f_L] > 0. \tag{6.2}$$

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

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

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

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

$$\hat{\beta}_n \in \underset{\beta_0 \in \Theta}{\operatorname{argmin}}\; s(y - X\beta_0)' X \Omega_n \big[s(y - X\beta_0), X\big] X' s(y - X\beta_0) \tag{6.3}$$

or

$$\hat{\beta}_n^{2S} \in \underset{\beta_0 \in \Theta}{\operatorname{argmin}}\; s(y - X\beta_0)' X \hat{\Omega}_n \big[s(y - X\hat{\beta}), X\big] X' s(y - X\beta_0), \tag{6.4}$$

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

We discuss Assumptions 6.1 - 6.7 and compare them with the ones required for LAD and quantile estimator consistency. On considering the special case where $X\Omega_n\big[s(y - X\beta_0), X\big]X' = I_n$ the identity matrix, the estimators in (6.3) - (6.4) coincide with the "quantile regression estimator" (with $\theta = 1/2$) studied by Fitzenberger (1997, Theorem 2.2). However, allowing for a weighting matrix different the identity matrix – as we do here – turns out to be important from the viewpoint of efficiency. *Stricto sensu*, the sign-based estimators in (6.3) - (6.4) and Fitzenberger (1997, Theorem 2.2) are not LAD estimators, because the size of residuals (through absolute values) do not appear in the objective function. This feature is crucial for relaxing assumptions on moments. The disturbances indeed appear in the objective function only through their sign transforms which possess finite moments at all orders. Consequently, no additional restriction need be imposed on the disturbance process (in addition to regularity conditions on the density). Only assumptions on the moments of x_t are used (see Assumption 6.2). There is very little work on LAD estimators properties with infinite variance errors; see Knight (1989) and Phillips (1991) who derive LAD asymptotic properties for an autoregressive model with infinite variance errors, which are in the domain of attraction of a stable law.

Assumption 6.1 on mixing is needed to apply a generic weak law of large numbers; see Andrews (1987) and White (2001). It was used by Fitzenberger (1997) with stationary linearly dependent processes. It covers, among other processes, stationary ARMA disturbances with continuously distributed innovations. Identification is provided by Assumptions 6.4 and 6.5. Assumption 6.5 is similar to Condition ID in Weiss (1991). Assumption 6.4 is usual in LAD estimator asymptotics.[8] It is analogous to Fitzenberger's (1997) conditions (ii.b) - (ii.c) and Weiss's (1991) CD condition. It implies that there is enough variation around zero to identify the median. This restricts the setup for some "bounded" heteroskedasticity in the disturbance process but not in the usual (variance-based) way. It is related to *diffusivity* $[2f(0)]^{-1}$, a dispersion measure adapted to median-unbiased estimators. Diffusivity indicates the vertical spread of a density rather than its horizontal spread, and appears in Cramér-Rao-type efficiency bounds for median-unbiased estimators; see Sung, Stangenhaus and David (1990) and So (1994). Assumption 6.6 implies that the weight matrix Ω_n is everywhere invertible, while Assumption 6.7 only requires local invertibility.

6.2. Asymptotic normality

Sign-based estimators are asymptotically normal under the following assumptions.

[8] 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* $, \forall n \in \mathbb{N}, \forall \lambda \in \mathbb{R},$

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

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

$$L(\beta) = \lim_{n\to\infty} \frac{1}{n} \sum_t \mathsf{E}\{x_t x_t' f_t[x_t'(\beta - \bar{\beta}) \,|\, x_1, \ldots, x_n]\} = \lim_{n\to\infty} L_n(\beta) \tag{6.5}$$

where

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

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

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

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

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

Then, we have the following result.

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

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

where $\hat{\beta}_n(\Omega_n)$ *is any estimator which minimizes* $D_S[\beta_0, \bar{\Omega}_n(\beta_0)]$ *in* (2.2),

$$S_n = [L_n(\bar{\beta})\Omega_n L_n(\bar{\beta})]^{-1} L_n(\bar{\beta})\Omega_n J_n \Omega_n L_n(\bar{\beta})[L_n(\bar{\beta})\Omega_n L_n(\bar{\beta})]^{-1}, \tag{6.8}$$

$$L_n(\bar{\beta}) = \frac{1}{n} \sum_t \mathsf{E}\big[x_t x_t' f_t\big(0 \,|\, x_1, \ldots, x_n\big)\big]. \tag{6.9}$$

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

$$[L_n(\bar{\beta}) \hat{J}_n^{-1} L_n(\bar{\beta})]^{-1/2} \sqrt{n} \big[\hat{\beta}_n(\hat{J}_n^{-1}) - \bar{\beta}\big] \xrightarrow{d} \mathrm{N}\big[0, I_p\big]. \tag{6.10}$$

This corresponds to the use of optimal instruments and quasi-efficient estimation. $\hat{\beta}_n(\hat{J}_n^{-1})$ has the same asymptotic covariance matrix as the LAD estimator. Thus, performance differences between the two estimators correspond to finite-sample features. This result contradicts the generally accepted idea that sign procedures involve a heavy loss of information. There is no loss induced by the use of signs instead of absolute values. Note again that we do not require that the disturbance process variance be finite. We only assume that the second-order moments of X are finite and the mixing property of $\{W_t, \ t = 1, \ldots\}$ holds.[9]

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

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

$$S_n = [L_n(\bar{\beta})]^{-1} \mathsf{E}\big[(1/n)\sum_{t=1}^{n} x_t x_t'\big] [L_n(\bar{\beta})]^{-1}. \tag{6.11}$$

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

$$S_n = \frac{1}{4f(0)^2}[\mathsf{E}(x_t x_t')]^{-1}. \tag{6.12}$$

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

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

Finally, we have two ways of making inference with signs: we can use the Monte Carlo (finite-sample) based method described in Coudin and Dufour (2009) and the classical asymptotic method. Let us list here the main differences between them. Monte Carlo inference relies on the pivotality of sign-based test statistics. The tests so obtained are valid (with controlled level) for any sample size if the mediangale Assumption 2.1 holds. When only the sign moment Assumption 2.2 holds, the Monte Carlo inference remains asymptotically valid. Asymptotic test levels are controlled. Besides, in simulations, Monte Carlo inference method appears to perform better in small samples than classical asymptotic methods, even if its use is only asymptotically justified [see Coudin and Dufour (2009)]. Nevertheless, this method has a drawback: its computational complexity. In contrast, classical asymptotic methods which yield tests with controlled asymptotic level under the sign moment Assumption 2.2 may be less time consuming. The choice between both is mainly a question of computational capacity. Classical asymptotic inference greatly relies on the way the asymptotic covariance matrix, which 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 sign-based estimators with the OLS and LAD estimators in terms of asymptotic bias and RMSE.

7.1. Simulation setup

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

$$y_t = x_t'\beta + u_t \quad t = 1, \ldots, n, \tag{7.1}$$

Table 1. Simulated models.

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

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

Cases A1 and A2 present *i.i.d.* normal observations without and with conditional heteroskedasticity. Case A3 involves weak nonlinear dependence in the error term. Case A4 presents a very unbalanced scheme in the design matrix (a case when the LAD estimator is known to perform badly). Cases B5, B6, B7 and B8 illustrate thick-tailed errors, heteroskedasticity and nonlinear dependence. Cases C9 to C11 illustrate different levels of autocorrelation in the error term with and without heteroskedasticity.

7.2. Bias and RMSE

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

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

For setups with strong heteroskedasticity and nonstationary disturbances (B5 - B8), we see that the sign-based estimators yield better results than both LAD and OLS estimators. Not far from the (optimal) LAD in case of Cauchy disturbances (B5), the sign-based estimators are the only ones which are reliable with nonstationary variance (B6 - B8).

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). However, when linear dependence is stronger (C11), this is no longer true. The *SHAC* sign-based estimator does not give better results than the non-corrected one in these examples.

To conclude, sign-based estimators are robust estimators less sensitive than the LAD estimator to unbalanced schemes in the explanatory variables and to heteroskedasticity. They are particularly adequate with heteroskedasticity or nonlinear dependence in the error term, even if the error term fails to be stationary. Finally, if the HAC correction improves the performance of test procedures in the presence of serial dependence [see Coudin and Dufour (2009)], it does not appear to do so for point estimation.

8. Empirical illustration

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

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

Table 2. Simulated bias and RMSE.

$n=50$		OLS		LAD		SF		SHAC	
$S=1000$		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
A1:	β_0	.003	.142*	.002	.179	.002	.179	.004	.178
	β_1	.003	.149*	.006	.184	.004	.182	.004	.182
	β_2	−.002	.149*	−.007	.186	−.006	.185	−.007	.183
	$\|\|\beta\|\|$	**.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*
	$\|\|\beta\|\|$	**.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
	$\|\|\beta\|\|$	**.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*
	$\|\|\beta\|\|$	**.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
	$\|\|\beta\|\|$	**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*
	$\|\|\beta\|\|$	**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
	$\|\|\beta\|\|$	**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
	$\|\|\beta\|\|$	$\mathbf{>10^{10}}$	$\mathbf{>10^{10}}$	$\mathbf{>10^{10}}$	$\mathbf{>10^{10}}$	**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*
	$\|\|\beta\|\|$	**.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*
	$\|\|\beta\|\|$	**.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
	$\|\|\beta\|\|$	**.070**	**1.29**	**.027**	**.430***	**.059**	**1.36**	**.054**	**1.37**

Note – $\|\cdot\|$ stands for the Euclidean norm. Best results for bias and RMSE are marked with a star (*).

Table 3. Regressions for personal income across U.S. States, 1880-1988: estimates of β

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

* Projection-based 95% CI.

** Asymptotic 95% CI.

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

$i = 1, \ldots, 48$, $T = 8$, 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 β-convergence at a rate of 2% per 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 suggests the presence of highly influential observations in all the periods but one; see Table 4 in the online appendix. Outliers are clearly identified in periods 1900-1920, 1940-1950, 1950-1960, 1970-1980, and 1980-1988.

Sign-based estimates are more stable than the least-squares ones. They vary between $[-.0147, .0364]$ whereas least-squares estimates vary between $[-.0149, .0431]$. This suggests that at least 12% of the least-squares estimates variability between sub-periods is due to the non-robustness of the least-squares method. 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 5(a) - 7(f) [see the Technical Appendix to this paper], 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 3 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 over the period 1930 - 1940 [see graphic 6(b)], whereas over the period 1980 - 1988, the same parameter is less precisely estimated and the p-value function leads to a wider confidence intervals [see graphic 7(f)].

9. Conclusion

In this paper, we have introduced inference tools which can be associated with the Monte Carlo based system presented in Coudin and Dufour (2009): the p-value function (and its individual projected versions) which gives a visual summary of all the inference available on a particular parameter, and Hodges-Lehmann-type sign-based estimators. The p-value function associates to each value of the parameter vector a confidence degree. It extends the confidence distribution concept to multidimensional parameters and relies on a reinterpretation of Fisher's fiducial distributions.

Parameter values least rejected by tests (given the sample realization and the sample size) constitute Hodges-Lehmann-type sign-based estimators. Those estimators are associated with the highest p-value. Hence, they are derived without referring to asymptotic conditions through the analogy principle. However, they turn out to be equivalent (in probability) to usual GMM estimators based on signs.

We then derived some general properties of sign-based estimators (invariance, median unbiasedness) and conditions under which consistency and asymptotic normality hold. In particular, we showed that sign-based estimators do require less assumptions on moment existence of the disturbances than usual LAD asymptotic theory. Simulation studies indicate that the proposed estimators are accurate in classical setups and more reliable than usual methods (LS, LAD) when heterogeneity or nonlinear dependence is present in the error term even in cases which 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.

References

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

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

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

Barnard, G. A. (1963), 'Comment on 'The spectral analysis of point processes' by M. S. Bartlett', *Journal of the Royal Statistical Society, Series B* **25**(2), 294.

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

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

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

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

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

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

Coudin, E. and Dufour, J.-M. (2009), ‘Finite-sample distribution-free inference in linear median regressions under heteroskedasticity and nonlinear dependence of unknown form’, *Econometrics Journal, 10th anniversary special edition* **12**(S1), S19–S49.

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

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

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

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

Dufour, J.-M. (2006), ‘Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in econometrics’, *Journal of Econometrics* **133**(2), 443–477.

Dufour, J.-M. and Jasiak, J. (2001), ‘Finite sample limited information inference methods for structural equations and models with generated regressors’, *International Economic Review* **42**(3), 815–843.

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

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

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

So, B. S. (1994), 'A sharp Cramér-Rao type lower-bound for median-unbiased estimators', *Journal of the Korean Statistical Society* **23**(1), 187–197.

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

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

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

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

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

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

by
Élise Coudin and Jean-Marie Dufour
INSEE-CREST and McGill University
November 2019

Technical appendix

A. Proofs

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

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

hence

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

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

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

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

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

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

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

$$\hat{\beta}(u, X, u) = -\hat{\beta}(-u, X, u) \tag{A.5}$$

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

$$u \sim -u \Rightarrow \hat{\beta}(u, X, u) \sim -\hat{\beta}(-u, X, u) \Rightarrow \mathrm{Med}\big(\hat{\beta}(u, X, u)\big) = 0. \tag{A.6}$$

Moreover, equation (5.5) implies:

$$\hat{\beta}(y, X, u) = \hat{\beta}(y - X\bar{\beta}, X, u) + \bar{\beta} = \hat{\beta}(u, X, u) + \bar{\beta}. \tag{A.7}$$

Finally, (A.6) and (A.7) entail $\mathrm{Med}[\hat{\beta}(y, X, u) - \bar{\beta}] = 0$. □

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

defined on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Set

$$\begin{aligned} q_t(W_t,\beta) &= \left[q_{t1}(W_t,\beta),\ldots,q_{tp}(W_t,\beta)\right]' \\ &= \left[s(y_t-x_t'\beta)x_{t1},\ldots,s(y_t-x_t'\beta)x_{tp}\right]',\ t=1,\ldots,n. \end{aligned} \tag{A.8}$$

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

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

$$\frac{1}{n}\sum_{t=1}^{n} q_{tk}(W_t,\beta) - \mathsf{E}[q_{tk}(W_t,\beta)] \xrightarrow{p} 0, \quad k=1,\ldots,p,\ \forall\beta\in\Theta. \tag{A.9}$$

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

$$q^H_{ik}(W_i,\beta,\rho) = \sup_{\hat\beta\in B(\beta,\rho)} q_{ik}(W_i,\hat\beta), \tag{A.10}$$

$$q_{Lik}(W_i,\beta,\rho) = \inf_{\hat\beta\in B(\beta,\rho)} q_{ik}(W_i,\hat\beta), \tag{A.11}$$

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

The last condition to check (A6) requires the following: let μ be a σ-finite measure which dominates each one of the marginal distributions of $W_t,\ t=1,2,\ldots$, and $p_t(w)$ 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 $\beta\in\Theta$, and

$$\int \sup_{t\geq 0,\ \beta\in\Theta} |q_{tk}(W,\ \beta)|p_t(w)d\mu(w) < \infty. \tag{A.12}$$

Since u_t is continuously distributed uniformly in t [Assumption 6.4(2)], we have $\mathbb{P}_t[u_t = x_t\beta] = 0,\ \forall\beta,$ uniformly in t. Then, q_{tk} is continuous in β everywhere except on a $\mathbb{P}_t$-negligible set. Finally, since q_{tk} is L_1-bounded and uniformly integrable, condition A6 holds. The generic law of large numbers (GWLLN) then implies:

$$\begin{array}{ll} \text{(a)} & \frac{1}{n}\sum_{i=0}^{n}\mathsf{E}[q_t(W_t,\ \beta)] \text{ is continuous on } \Theta \text{ uniformly over } n\geq 1, \\ \text{(b)} & \sup_{\beta\in\Theta}\left|\frac{1}{n}\sum_{t=0}^{n} q_t(W_t,\ \beta) - \mathsf{E}q_t(W_t,\ \beta)\right| \to 0 \quad \text{as } n\to\infty \text{ in probability under } \mathbb{P}. \end{array} \tag{A.13}$$

The **Consistency Theorem** is 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 differs slightly. For simplicity, the true value is taken to be 0. First, the generic law of large numbers implies that

$$\lim_{n\to\infty} \frac{1}{n} \sum_t \mathsf{E}[s(u_t - x_t'\beta)x_{tk}] \text{ is continuous on } \Theta,\, k = 1, \ldots, p. \tag{A.14}$$

Define

$$Q_n^k(\beta) = \frac{1}{n}\Big|\sum_{t=1}^{n} x_{kt} s(u_t - x_t'\beta)\Big|,\ k = 1, \ldots, p, \tag{A.15}$$

$$Q_n^{Ek}(\beta) = \frac{1}{n}\Big|\sum_{t=1}^{n} \mathsf{E}[x_{kt} s(u_t - x_t'\beta)]\Big|,\ k = 1, \ldots, p, \tag{A.16}$$

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

$$\frac{1}{n^2}\sum_{k=1}^{p} \big[\sum_t x_{kt} s(u_t - x_t'\beta)\big]^2 = \sum_k [Q_n^k(\beta)]^2. \tag{A.17}$$

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

$$\mathbb{P}\big[\sum_k [Q_n^k(\beta_n)]^2 < \sum_k [Q_n^k(0)]^2 + \varepsilon/3\big] \geq 1 - \delta. \tag{A.18}$$

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

$$[Q_n^{Ek}(\beta_n)]^2 < [Q_n^k(\beta_n)]^2 + (\varepsilon/3p),\ k = 1, \ldots, p,\ \text{with probability approaching one as } n \to \infty, \tag{A.19}$$

hence,

$$\sum_k [Q_n^{Ek}(\beta_n)]^2 < \sum_k [Q_n^k(\beta_n)]^2 + \varepsilon/3,\ \text{with probability approaching one as } n \to \infty. \tag{A.20}$$

Using the same argument at $\beta = 0$, we have

$$\sum_k [Q_n^k(0)]^2 < \sum_k [Q_n^{Ek}(0)]^2 + \varepsilon/3,\ \text{with probability approaching one as } n \to \infty. \tag{A.21}$$

By (A.20), (A.18) and (A.21), this in turn implies:

$$\sum_k [Q_n^{Ek}(\beta_n)]^2 < \sum_k [Q_n^{Ek}(0)]^2 + \varepsilon,\ \text{with probability approaching one as } n \to \infty. \tag{A.22}$$

This holds for any ε, with probability approaching one. Let $\mathbf{N}$ be any open subset of Θ containing 0. Since $\Theta \cap \mathbf{N}^c$ is compact and $\lim_n \sum_k [Q_n^{*k}(\beta)]^2$ is continuous [see (A.14],

$$\exists \beta^* \in \Theta \cap\ \mathbf{N}^c \text{ such that } \sup_{\beta\in\Theta\cap\mathbf{N}^c} \lim_n \sum_k [Q_n^{Ek}(\beta)]^2 = \lim_n \sum_k [Q_n^{Ek}(\beta^*)]^2. \tag{A.23}$$

Provided 0 is the unique minimizer, we have:

$$\lim_n \sum_k [Q_n^{Ek}(\beta^*)]^2 > \lim_n \sum_k [Q_n^{Ek}(0)]^2,\ \text{with probability one .} \tag{A.24}$$

Hence, setting

$$\varepsilon = \frac{1}{2}\left\{\lim_n \sum_k [Q_n^{Ek}(\beta^*)]^2\right\}, \tag{A.25}$$

it follows that, with probability close to one,

$$\lim_n \sum_k [Q_n^{Ek}(\beta_n)]^2 < \frac{1}{2}\big[\lim_n \sum_k [Q_n^{Ek}(\beta^*)]^2 + \lim_n \sum_k [Q_n^{Ek}(0)]^2\big] < \sup_{\beta \in \Theta \cap \mathbf{N}^c} \lim_n \sum_k [Q_n^{Ek}(\beta)]^2. \tag{A.26}$$

Hence, $\beta_n \in \mathbf{N}$. As this holds for any open subset $\mathbf{N}$ of Θ, we conclude that β_n converges to 0.
For **identification**, the uniqueness of the minimizer of the sign-objective function is insured by the set of identification conditions 2.2, 6.4, 6.5, and 6.6. These conditions and consequently the proof, are close to those of Weiss (1991) and Fitzenberger (1997) for the LAD and quantile estimators. We wish to show that the limit problem does not admit another solution. When $\bar{\Omega}_n(\beta)$ defines a norm for each β (Assumption 6.6), this assertion is equivalent to

$$\lim_{n\to\infty} \mathsf{E}\big[\frac{1}{n}\sum_t s(u_t - x_t'\delta)x_t\big] = 0 \Rightarrow \delta = 0,\ \ \delta \in \mathbb{R}^p, \tag{A.27}$$

and

$$\lim_{n\to\infty} \big|\mathsf{E}\big[\frac{1}{n}\sum_t s(u_t - x_t'\delta)x_t'\delta\big]\big| = 0 \Rightarrow \delta = 0,\ \ \delta \in \mathbb{R}^p. \tag{A.28}$$

Let

$$A(\delta) = \mathsf{E}[\frac{1}{n}\sum_t s(u_t - x_t'\delta)x_t | x_1, \ldots, x_n]. \tag{A.29}$$

Then,

$$\mathsf{E}[A(\delta)] = \mathsf{E}\big[\frac{1}{n}\sum_t s(u_t - x_t'\delta)x_t\big] = \mathsf{E}\{\mathsf{E}\big[\frac{1}{n}\sum_t s(u_t - x_t'\delta)x_t | x_1, \ldots, x_n\big]\}. \tag{A.30}$$

Note that

$$\mathsf{E}[s(u_t - x_t'\delta)|x_1, \ldots, x_n] = 2\left[\frac{1}{2} - \int_{-\infty}^{x_t'\delta} f_t(u|x_1, \ldots, x_n)du\right] = -2\int_0^{x_t'\delta} f_t(u|x_1, \ldots, x_n)du] \tag{A.31}$$

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

$$\begin{aligned} A(\delta) \;\; = \;\; & \frac{2}{n}\sum x_t'\delta\Big\{I_{\{|x_t'\delta|>\tau\}}\Big[I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta} -f_t(u|x_1, \ldots, x_n)du + I_{\{x_t'\delta\le 0\}}\int_{x_t'\delta}^0 f_t(u|x_1, \ldots, x_n)du\Big] \\ & + I_{\{|x_t'\delta|\le\tau\}}\Big[I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta} -f_t(u|x_1, \ldots, x_n)du + I_{\{x_t'\delta\le 0\}}\int_{x_t'\delta}^0 f_t(u|x_1, \ldots, x_n)du\Big]\Big\}, \end{aligned} \tag{A.32}$$

hence

$$\begin{aligned} \mathsf{E}[A(\delta)] = \mathsf{E}\Big\{\frac{2}{n}\sum x_t'\delta\Big[I_{\{|x_t'\delta|>\tau\}}\Big(I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta} -f_t(u|x_1, \ldots, x_n)du + I_{\{x_t'\delta\le 0\}}\int_{x_t'\delta}^0 f_t(u|x_1, \ldots, x_n)du\Big) \\ + I_{\{|x_t'\delta|\le\tau\}}\Big(I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta} -f_t(u|x_1, \ldots, x_n)du + I_{\{x_t'\delta\le 0\}}\int_{x_t'\delta}^0 f_t(u|x_1, \ldots, x_n)du\Big)\Big]\Big\}. \end{aligned} \tag{A.33}$$

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

$$
\begin{aligned}
|\mathsf{E}(A)| &= \mathsf{E}\big[\frac{2}{n}\sum x_t'\delta I_{\{|x_t'\delta|>\tau\}}\big(I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta} f_t(u|x_1,\ldots,x_n)du - I_{\{x_t'\delta\le 0\}}\int_{x_t'\delta}^0 f_t(u|x_1,\ldots,x_n)du\big)\big] \\
&\quad +\mathsf{E}\big[\frac{2}{n}\sum x_t'\delta I_{\{|x_t'\delta|\le\tau\}}\big(I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta} f_t(u|x_1,\ldots,x_n)du - I_{\{x_t'\delta\le 0\}}\int_{x_t'\delta}^0 f_t(u|x_1,\ldots,x_n)du\big)\big] \\
&\ge \mathsf{E}\big[\frac{2}{n}\sum I_{\{|x_t'\delta|>\tau\}}\big(x_t'\delta I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta} f_t(u|x_1,\ldots,x_n)du \\
&\qquad -x_t'\delta I_{\{x_t'\delta\le 0\}}\int_{x_t'\delta}^0 f_t(u|x_1,\ldots,x_n)du\big)\big] \qquad \text{(A.34)} \\
&\ge \mathsf{E}\big\{\frac{2}{n}\sum I_{\{|x_t'\delta|>\tau\}}\big[x_t'\delta I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta} f_t(u|x_1,\ldots,x_n)du \\
&\qquad -x_t'\delta I_{\{x_t'\delta\le 0\}}\int_{x_t'\delta}^0 f_t(u|x_1,\ldots,x_n)du\big][f_t(0|x_1,\ldots,x_n)>f_L]p_1\big\} \qquad \text{(A.35)} \\
&\ge p_1\mathsf{E}\big\{\frac{2}{n}\sum I_{\{|x_t'\delta|>\tau\}}\tau f_L d|f_t(0|x_1,\ldots,x_n)>f_L\big\} \qquad \text{(A.36)} \\
&\ge \tau p_1 f_L d\frac{2}{n}\sum \mathbb{P}[|x_t'\delta|>\tau|f_t(0|x_1,\ldots,x_n)>f_L)]. \qquad \text{(A.37)}
\end{aligned}
$$

To obtain inequation (A.34), we just note that each term is positive. For the inequation (A.35) we use Assumption 6.4. For inequation (A.36) we minorate $|x_t'\delta|$ 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, that

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

hence, we conclude on the uniqueness of the minimum, which was the last step to ensure 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 (6.5), is the derivative of $\lim_{n\to\infty}\frac{1}{n}\sum_t \mathsf{E}\big[s\big(u_t - x_t'(\beta-\bar\beta)\big)x_t\big]$. Then, we check the conditions for applying Theorem 7.2 of Newey and McFadden (1994).

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

$$\frac{1}{n}\sum_{t=0}^{n}\mathsf{E}\big[s\big(u_t - x_t'(\beta-\bar\beta)\big)x_t\big] \qquad \text{(A.38)}$$

is continuous on Θ uniformly over n. Moreover Assumption 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 (Assumption 6.8), dominated convergence allows us to write:

$$\frac{\partial}{\partial\beta}\mathsf{E}\big[x_t s\big(u_t - x_t'(\beta-\bar\beta)\big)\big] = \mathsf{E}\big[x_t x_t' f_t\big(x_t'(\beta-\bar\beta)|x_1,\ldots,x_n\big)\big] \qquad \text{(A.39)}$$

These conditions imply:

$$L_n(\beta) = \frac{1}{n}\sum_{t=1}^{n} \mathsf{E}\left[x_t x_t' f_t\left(x_t'(\beta-\bar{\beta})|x_1, \ldots, x_n\right)\right] \underset{n\to\infty}{\to} L_n(\beta) \tag{A.40}$$

uniformly in β. Uniform convergence implies that $\lim_n \frac{1}{n}\sum_{t=0}^{n} \mathsf{E}\left[s\left(u_t - x_t'(\beta-\bar{\beta})\right)x_t\right]$ is differentiable with derivative $L(\beta)$.

We now apply Theorem 7.2 of Newey and McFadden (1994) which presents asymptotic normality of a minimum distance consistent estimator with nonsmooth objective function and weight matrix $\Omega_n \xrightarrow{p} \Omega$ symmetric positive definite. Thus, under conditions for consistency (2.2, 6.1-6.6), we check that the following conditions hold:

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

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

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

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

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

Condition (i) is fulfilled by the moment Assumption 2.2. Condition (ii) is fulfilled by the first part of our proof and Assumption 6.10. Then, Condition (iii) is implied by 6.3. Using the mixing Assumption 6.9 of $\{u_t, X_t\}_{t=1,2,\ldots}$ and conditions 2.2, 6.2, 6.7 and 6.11, we apply a White-Domowitz central limit theorem [see White (2001), Theorem 5.20]. This fulfills condition (iv) of Theorem 7.2 in Newey and McFadden (1994):

$$\sqrt{n}J_n^{-1/2}Q_n(\bar{\beta}) \to \mathrm{N}[0, I_p] \tag{A.41}$$

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

$$\sqrt{n}S_n^{-1/2}\left[\hat{\beta}(\Omega_n) - \bar{\beta}\right] \to \mathrm{N}[0, I_p]. \tag{A.42}$$

The asymptotic covariance matrix S is given by the limit of

$$S_n = [L_n(\bar{\beta})\Omega_n(\bar{\beta})L_n(\bar{\beta})]^{-1}L_n(\bar{\beta})\Omega_n(\bar{\beta})J_n\Omega_n(\bar{\beta})L_n(\bar{\beta})[L_n(\bar{\beta})\Omega_n(\bar{\beta})L_n(\bar{\beta})]^{-1}. \tag{A.43}$$

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

$$\sqrt{n}S_n^{-1/2}\left[\hat{\beta}(\hat{J}_n^{-1}) - \bar{\beta}\right] \to \mathrm{N}[0, I_p] \tag{A.44}$$

with

$$S_n = [L_n(\bar{\beta})\hat{J}_n^{-1}L_n(\bar{\beta})]^{-1}. \tag{A.45}$$

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

Table 4. Summary of regression diagnostics.

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

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

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

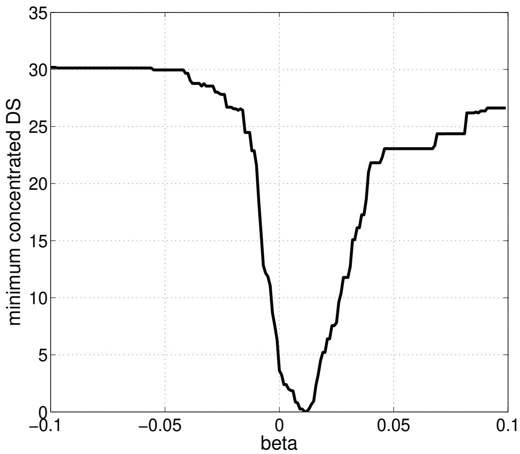
$$\lim_{n\to\infty} S_n = \lim_{n\to\infty} \frac{n^2}{4}\{\sum_t \mathsf{E}\big[x_t x_t' f_t(0|X)\big]\}^{-1}\mathsf{E}(x_t x_t')\{\sum_i \mathsf{E}\big[x_t x_t' f_t(0|X)\big]\}^{-1}. \tag{A.46}$$

□

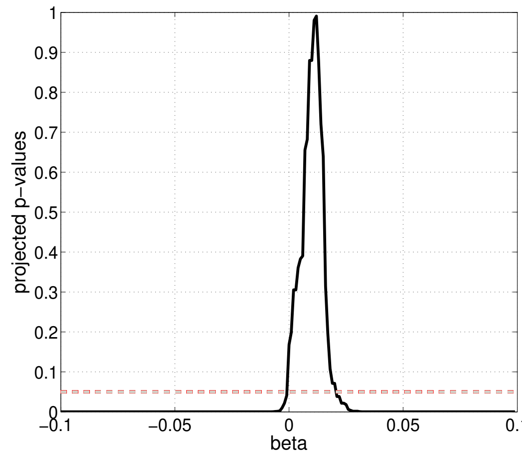
B. Convergence data: concentrated statistics and p-values

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

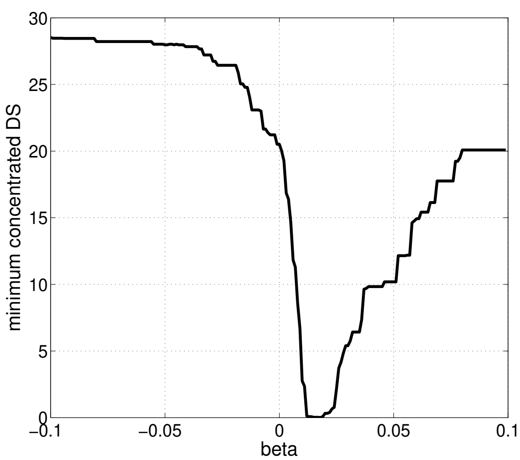
Figure 5. Concentrated statistics and projected *p*-values (1880-1930)



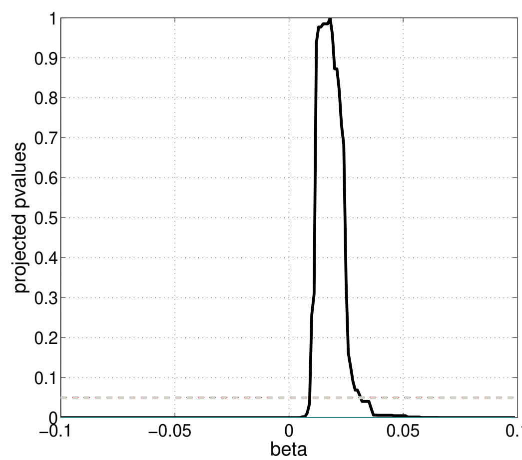
(a) Basic equation: 1880-1900: concentrated DS



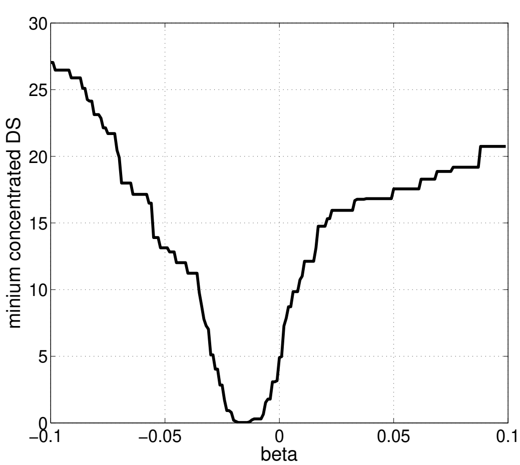
(b) Basic equation: 1880-1900: projected *p*-value



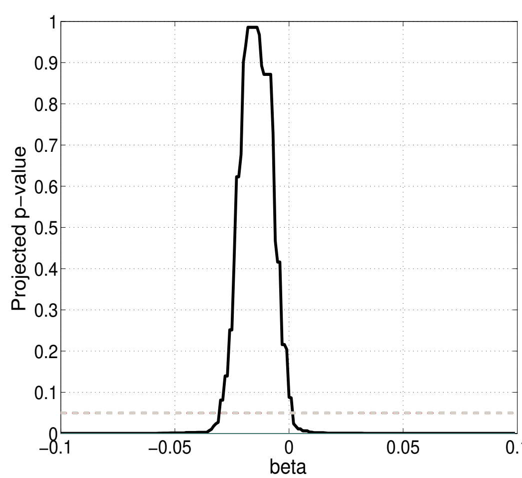
(c) Basic equation: 1900-20: concentrated DS



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

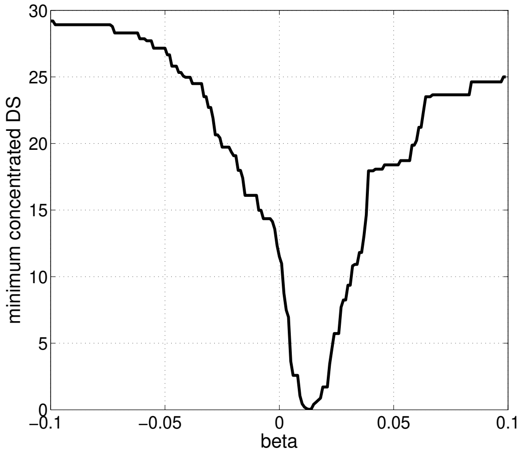


(e) Basic equation: 1920-30: concentrated DS

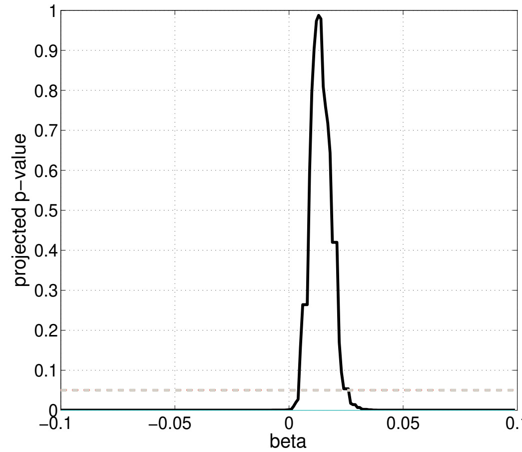


(f) Basic equation: 1920-30: projected *p*-value

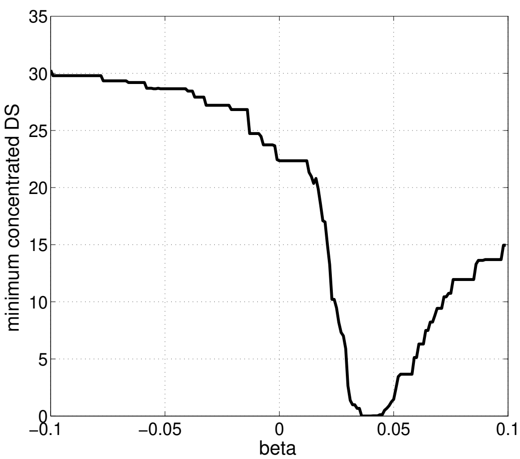
Figure 6. Concentrated statistics and projected *p*-values (1930-1960)



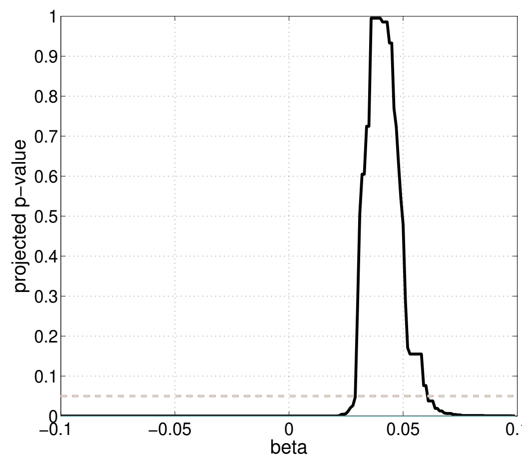
(a) Basic equation: 1930-40: concentrated DS



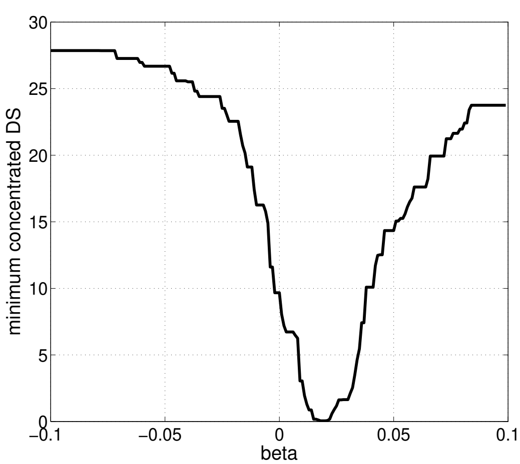
(b) Basic equation: 1930-40: projected *p*-value



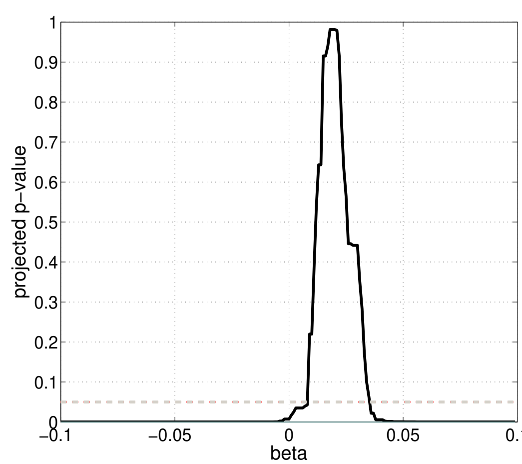
(c) Basic equation: 1940-50: concentrated DS



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

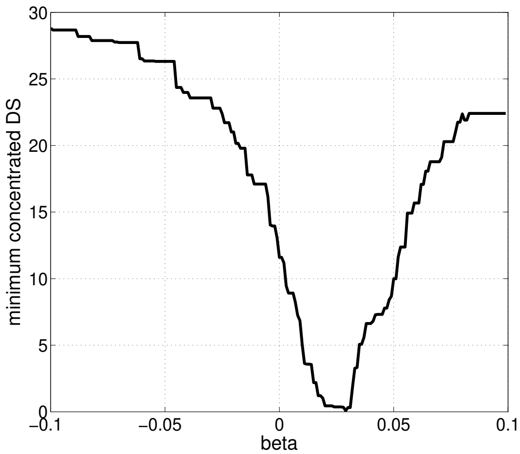


(e) Basic equation: 1950-60: concentrated DS

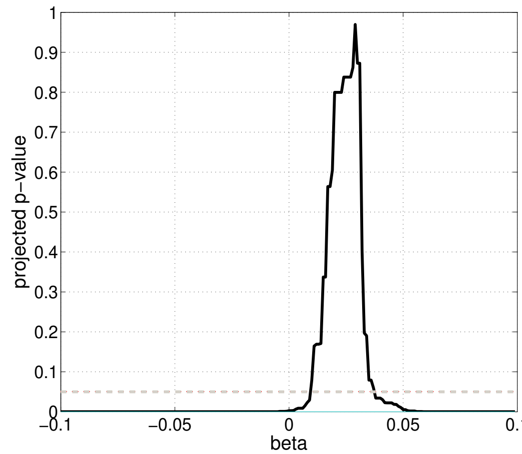


(f) Basic equation: 1950-60: projected *p*-value

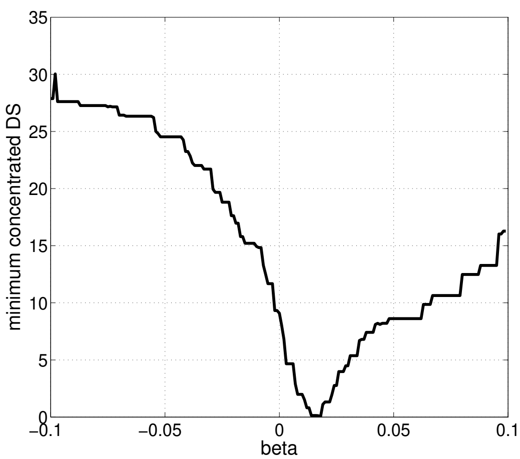
Figure 7. Concentrated statistics and projected *p*-values (1960-1988)



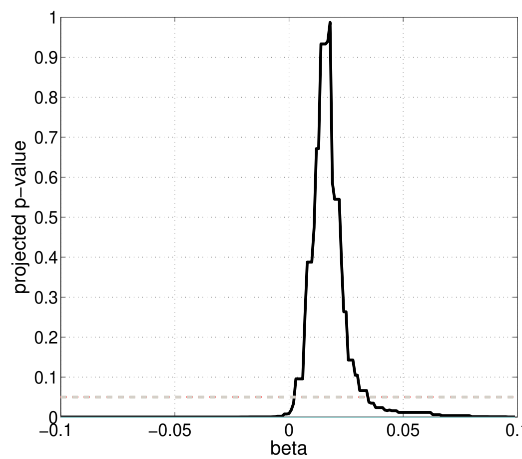
(a) Basic equation: 1960-70: concentrated DS



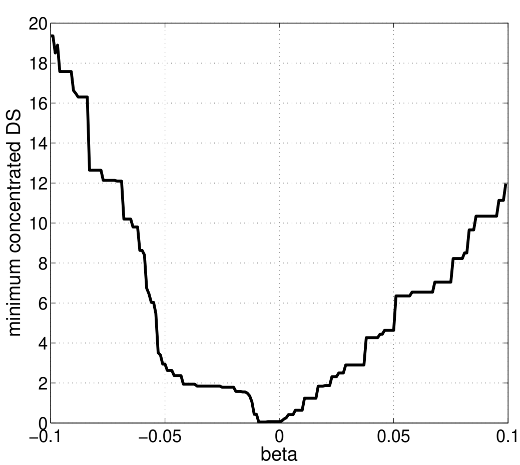
(b) Basic equation: 1960-70: projected *p*-value



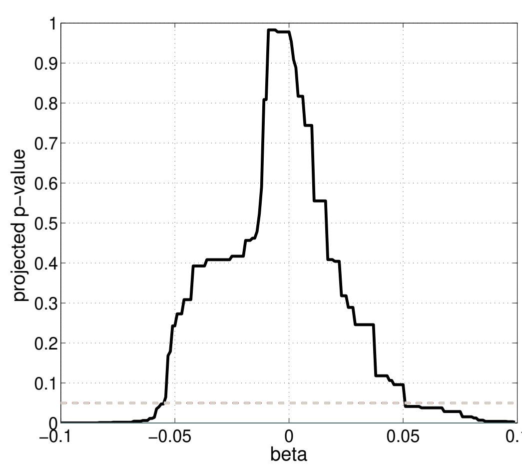
(c) Basic equation: 1970-80: concentrated DS



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



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



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