**2011s-24**

# Robust Sign-Based and Hodges-Lehmann Estimators in Linear Median Regressions with Heterogenous Serially Dependent Errors

*Elise Coudin, Jean-Marie Dufour*

**Série Scientifique**
***Scientific Series***

**Montréal**
**Février 2011**



CIRANO
Allier savoir et décision

Centre interuniversitaire de recherche en analyse des organisations

**ISSN 1198-8177**

**Partenaire financier**

Développement économique, Innovation et Exportation
Québec

# Robust Sign-Based and Hodges-Lehmann Estimators in Linear Median Regressions with Heterogenous Serially Dependent Errors[*]

*Elise Coudin*[†], *Jean-Marie Dufour*[‡]

### *Abstract*

We propose estimators for 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 or nonlinear) including GARCH-type and stochastic volatility of unknown order. The estimators follow from a reverse inference approach, based on the class of distribution-free sign tests proposed in Coudin and Dufour (2009, *Econometrics J.*) under a mediangale assumption. As a result, the estimators inherit strong robustness properties from their generating tests. Since the proposed estimators are based on maximizing a test statistic (or a p-value function) over different null hypotheses, they can be interpreted as Hodges-Lehmann-type (HL) estimators. It is easy to adapt the sign-based estimators to account for linear serial dependence. Both finite-sample and large-sample properties are established under weak regularity conditions. The proposed estimators are median unbiased (under symmetry and estimator unicity) and satisfy natural equivariance properties. Consistency and asymptotic normality are established without any condition on error moment existence, allowing for heterogeneity (or heteroskedasticity) of unknown form, noncontinuous distributions, and very general serial dependence (linear or nonlinear). These conditions are considerably weaker than those used to show corresponding results for LAD estimators. In a Monte Carlo study on bias and mean square error, we find that sign-based estimators perform better than LAD-type estimators, especially in heteroskedastic settings. The proposed procedures are applied to a trend model of the Standard and Poor's composite price index, where disturbances are affected by both heavy tails (non-normality) and heteroskedasticity.

**Key words** sign test, median regression, Hodges-Lehmann estimator, p-value; least absolute deviations, quantile regression; simultaneous inference, Monte Carlo tests, projection methods, nonnormality, heteroskedasticity; serial dependence; GARCH; stochastic volatility.

**Codes JEL** : C13, C12, C14, C15.

*Running head: Robust and HL sign-based estimators*

---

[*] The authors thank Magali Beffy, Marine Carrasco, Frédéric Jouneau, Marc Hallin, Thierry Magnac, Bill McCausland, Benoit Perron, and Alain Trognon for useful comments and constructive discussions. Earlier versions of this paper were presented at the 2005 Econometric Society World Congress (London), the Econometric Society European Meeting 2007 (Budapest), CREST (Paris), the Conference on Nonparametric statistics and time series in honor of Marc Hallin (Bruxelles, December 2009), and the First French Econometrics Conference celebrating Alain Monfort (Toulouse, December 2009). This work was supported by theWilliam Dow Chair in Political Economy (McGill University), the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Bank of Canada (Research Fellowship), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Institut de finance mathématique de Montréal (IFM2), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds de recherche sur la société et la culture (Québec).

[†] Centre de recherche en économie et statistique (CREST-ENSAE). Mailing address: Laboratoire de micro économetrie, CREST, Timbre J390, 15 Bd G. Péri, 92245 Malakoff Cedex, France. E-mail: elise.coudin@ensae.fr.
TEL: 33 (0) 1 41 17 77 33; FAX: 33 (0) 1 41 17 76 34.

[‡] Corresponding author. 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 519, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 8879; FAX: (1) 514 398 4938; e-mail: jean-marie.dufour@mcgill.ca.
Webpage: http://www.jeanmariedufour.com

# Contents

## List of Definitions, Propositions and Theorems



## List of Tables

# 1. Introduction

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

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

Specifically, we study the problem of estimating the parameters of a linear median regression without any assumption on the shape of the error distribution – including no condition on the existence of moments at any order – allowing for heterogeneity (or heteroskedasticity) of unknown form (including GARCH-type dependence and stochastic volatility of unknown order), noncontinuous distributions, and very general serial dependence. We adopt a *reverse inference approach* based on the distribution-free tests proposed in Coudin and Dufour (2009). The test statistics are quadratic forms of the constrained signs (aligned with respect to the null hypothesis) with a weighting matrix that may also depend on the constrained signs. The null distributions of these statistics remain the same under a wide set of distributional assumptions on model errors (as described above). We propose to estimate the parameters of the median regression by minimizing these sign-based test statistics over different null hypotheses. Since the tests used to generate them are remarkably robust, the estimators inherit strong robustness properties.

The proposed estimators can be viewed as GMM estimators based on a nondifferentiable objective function originally derived as a distribution-free test statistic. This feature also means that the distribution of the criterion function is completely known under a wide array of nonparametrically specified data generating processes, as opposed to setups where only the mean of the estimating function is set (the moment equations). Since the estimators are based on maximizing a test statistic over different null hypotheses, they can also be interpreted as Hodges-Lehmann-type (HL) estimators [Hodges and Lehmann (1963)]. When the test statistic is pivotal (*i.e.*, the null distribution is the same irrespective of the value set by the null hypothesis), the estimator also maximizes the $p$-value associated with different tested parameter values. In other words, if the null hypothesis has the form $H_0(\beta_0) : \beta = \beta_0$, the estimator corresponds to the value of $\beta_0$ which is "least rejected" by the test (*i.e.*, has the highest $p$-value).[1]

Both finite-sample and large-sample properties of sign-based estimators are established under weak regularity conditions. We show they are median unbiased (under symmetry and estimator unicity) and possess equivariance properties with respect to linear transformations of model variables. Consistency and asymptotic normality are established without any moment existence assumption on the errors, allowing noncontinuous distributions, heterogeneity and general serial dependence of unknown form. These conditions are considerably weaker than those usually used to obtain corresponding results for LAD estimators; see Bassett and Koenker (1978), Bloomfield and Steiger (1983), Powell (1984), Phillips (1991), Pollard (1991), Weiss (1991), Fitzenberger (1997), Knight (1998), El Bantli and Hallin (1999) and the references therein. In particular, asymptotic normality and consistency hold for heavy-tailed disturbances which may not have finite variances. This interesting property is induced by the sign transformation. Signs of residuals always possess finite moments, so no further restriction on the disturbance moments is required. Except for Knight (1989) and Phillips (1991), who considered the case of autoregressive models, the distribution of LAD estimators in regressions where the error variances may not exist has received little attention. In general, LAD estimators and the sign-based estimators proposed here follow from different optimization rules, and they can be quite different.

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

[1] Hodges and Lehmann (1963) proposed this general principle to obtain an estimate of a location parameter from rank tests. For some extensions to regressions with i.i.d. errors, see Jureckova (1971), Jaeckel (1972), and Koul (1971).

and, in presence of linear dependence, a HAC-type correction for heteroskedasticity and autocorrelation can be included in the criterion function.

The construction of sign-based estimators as Hodges-Lehmann estimators makes these a natural complement of the finite-sample tests used to generate them. The latter rely on the exact distribution of the corresponding sign-based test statistics, do not involve nuisance parameters, and allow one to control test levels in finite samples under heteroskedasticity and nonlinear dependence of unknown form. In Coudin and Dufour (2009), Monte Carlo test methods [Dwass (1957), Barnard (1963) and Dufour (2006)] are combined with test inversion and projection techniques [Dufour (1990, 1997), Dufour and Kiviet (1998), Abdelkhalek and Dufour (1998), Dufour and Jasiak (2001), Dufour and Taamouti (2005)] to build confidence sets and test general hypotheses.[2] There is no need to estimate the error density at zero in contrast with tests that rely on kernel estimates of the LAD asymptotic covariance matrix.[3] Furthermore, when the test criteria are modified to cover linear dependence, the resulting inference is asymptotically valid. The conjunction of sign-based tests, projection-based confidence regions, and sign-based estimators thus provides a complete system of inference, which is valid for any given sample size under very weak distributional assumptions and remains asymptotically valid under even weaker conditions (including allowance for linear dependence in regression disturbances).

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

Finally, we present an empirical application to financial data. We study a trend model for the Standard and Poor's Composite Price Index, over the period 1928-1987 as well as the 1929 crash period (which is characterized by huge price volatilities). The data are affected by serial dependence, heavy tails (non-normality) and heteroskedasticity.

The paper is organized as follows. Section 2 presents the model and the class of tests we exploit. In section 3, we define the proposed family of sign-based estimators. The finite-sample properties of the sign-based estimators are studied in section 4, while their asymptotic properties are considered

[2] 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).

[3] In the *i.i.d.* error case, Honore and Hu (2004) observed in simulations that kernel-based estimates of the asymptotic standard error of the median-based estimator tend to be too small, so the associated tests tend to overreject the null hypothesis. Other estimates of the LAD asymptotic covariance matrix can be obtained by bootstrap procedures [design matrix bootstrap in Buchinsky (1995, 1998), block bootstrap in Fitzenberger (1997), Bayesian bootstrap in Hahn (1997)] and resampling methods [Parzen, Wei and Ying (1994)]. But the justification of these also rely on usual asymptotic regularity conditions.

in section 5. In section 6, we present the results of our simulation study of bias and RMSE. The empirical application is reported in section 7. We conclude in section 8. Appendix A contains the proofs.

## 2. Framework

We will now summarize the general framework we study and define the test statistics on which the estimation methods we propose are based.

### 2.1. Model

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

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

where $y_t$ is a dependent variable, $x_t = (x_{t1}, \ldots, x_{tp})'$ is a $p$-vector of explanatory variables, and $u_t$ is an error process. The $x_t$'s may be random or fixed. In the sequel, $y = (y_1, \ldots, y_n)' \in \mathbb{R}^n$ will denote the dependent variable vector, $X = (x_1, \ldots, x_n)' \in \mathbb{R}^{n\times p}$ the $n \times p$ matrix of explanatory variables, and $u = (u_1, \ldots, u_n)' \in \mathbb{R}^n$ the disturbance vector. Moreover, $F_t(\,\cdot\,|x_1, \ldots, x_n)$ represents the distribution function of $u_t$ conditional on $X$. This framework is also used in Coudin and Dufour (2009).

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

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

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

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

**Assumption 2.1** WEAK CONDITIONAL MEDIANGALE. *Let* $\mathcal{F}_t = \sigma(u_1, \ldots, u_t, X)$, *for* $t \geq 1$.

$\mathbf{u}$ *in the adapted sequence* $\mathcal{S}(\mathbf{u}, \mathcal{F})$ *is a weak mediangale conditional on* $X$ *with respect to* $\{\mathcal{F}_t : t = 1, 2, \dots\}$ *iff* $\mathsf{P}[u_1 < 0|X] = \mathsf{P}[u_1 > 0|X]$ *and*

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

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

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

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

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

### 2.2. Quadratic sign-based tests

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

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

where $\bar{\Omega}_n(\beta_0) = \Omega_n\big[s(y - X\beta_0), X\big]$ is a $p \times p$ positive definite weight matrix which may depend on the constrained signs. If the disturbances follow a weak mediangale (Assumption 2.1), sign-based statistics of this form constitute pivotal functions: the distribution of $D_S[\beta_0, \bar{\Omega}_n(\beta_0)]$ conditional on $X$ is completely determined under $H_0(\beta_0)$ and can be simulated; see Coudin and Dufour (2009). Even though the distribution of $D_S[\beta_0, \bar{\Omega}_n(\beta_0)]$ depends on $X$ and $\Omega_n\big[\,\cdot\,\big]$ under $H_0(\beta_0)$, critical values can be approximated to any degree of precision by simulation. Alternatively, exact Monte Carlo tests can be built using a randomized tie correction procedure [Dufour (2006)]. So we can get an exact test of $H_0(\beta_0)$. The fact that $\Omega_n\big[\,\cdot\,\big]$ depends on the data only through $s(y - X\beta_0)$ plays a central role in generating this feature.

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

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

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

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

Note also that the confidence region

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

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

## 3. Robust and Hodges-Lehmann sign-based estimators

We will now exploit the tests described in the previous section to derive robust estimators of $\beta$. We first define the estimates and then discuss their interpretation as Hodges-Lehmann estimators.

### 3.1. Sign-based estimators

In view of the above distributional properties, we consider estimators $\hat{\beta}_n = \hat{\beta}_n(y,\, X,\, D_S)$ obtained by minimizing the sign statistic $D_S[\beta_0,\, \bar{\Omega}_n(\beta_0)]$ :

$$D_S[\hat{\beta}_n,\, \bar{\Omega}_n(\hat{\beta}_n)] = \min_{\beta_0 \in \Theta} D_S[\beta_0,\, \bar{\Omega}_n(\beta_0)] \tag{3.1}$$

where $\Theta$ is a subset of $\mathbb{R}^p$ (for example, an appropriate compact set). This family of estimators includes as special cases estimators already studied in the literature in the context of *i.i.d.* errors. Namely, the sign-based estimators proposed by Boldin et al. (1997) can be obtained by taking $\Omega_n = I_p$ or $\Omega_n = (X'X)^{-1}$ :

$$SB(\beta_0) \equiv D_S[\beta_0,\, I_p] = s(y - X\beta_0)'XX's(y - X\beta_0) \equiv SB(\beta_0)\,, \tag{3.2}$$

$$SF(\beta_0) \equiv D_S\big[\beta_0,\, (X'X)^{-1}\big] = s(y - X\beta_0)'X(X'X)^{-1}X'\, s(y - X\beta_0)\,. \tag{3.3}$$

Such estimators can be interpreted as GMM estimators based on the moment condition $\mathsf{E}[X's(y - X\beta_0)] = 0$. This condition has the special feature that the estimating function $X's(y - X\beta)$ is not differentiable with respect to $\beta$, while its distribution is completely determined in a general nonparametric setup.

Since the function $D_S[\beta_0,\, \bar{\Omega}_n(\beta_0)]$ is non-negative and can only take a finite number of values (signs are limited to the three distinct values $-1,\, 0,\, 1$), problem (3.1) always possesses at least one solution. Further, if $\Omega_n\big[s(y - X\beta_0),\, X\big]$ is continuous with respect to $s(\cdot)$, $D_S[\beta_0,\, \bar{\Omega}_n(\beta_0)]$ is continuous almost everywhere (with respect to the Lebesgue measure), the existence of a bounded solution can be guaranteed by restricting $\beta_0$ to a compact subset $\Theta \subseteq \mathbb{R}^p$ [for example, see Assumption 5.3 below]. Clearly, the solution may not be unique, and there is a set

$$M(y,\, X) \equiv \operatorname*{arg\,min}_{\beta_0 \in \Theta} D_S[\beta_0,\, \bar{\Omega}_n(\beta_0)] \tag{3.4}$$

of possible solutions. To get a unique solution, one may add a choice criterion, such as minimizing an appropriate norm or distance among the minimizers of the objective function.[4] Minimizing $D_S[\beta_0,\, \bar{\Omega}_n(\beta_0)]$ is a nonlinear problem and no general closed-form analytical solution is available. Further, the function is discrete and not (everywhere) differentiable. So we need to use nonlinear optimization algorithm that can handle such functions, such as the simplex algorithm or simulated annealing; see Goffe, Ferrier and Rogers (1994) and Press, Teukolsky, Vetterling and Flannery (2002).[5]

---

[4] In general, a unique solution may always be selected by virtue of the axiom of choice.

[5] For further discussion of estimation based on a non-smooth criterion, see Honoré and Powell (1994), Boldin et al.

In order to allow for dependence not covered by the mediangale assumption (2.2), such as an ARMA structure in $u_t$, we can consider sign-based statistics where the weighting matrix is the inverse of an HAC-type covariance matrix estimator:

$$D_S\big[\beta_0,\, \bar{J}_n(\beta_0)^{-1}\big] = s(y - X\beta_0)'X\big(\hat{J}_n[s(y - X\beta_0),\, X]\big)^{-1}X's(y - X\beta_0) \tag{3.5}$$

where $\bar{J}_n(\beta_0) = \hat{J}_n[s(y-X\beta_0),\, X]$ accounts for the dependence among the signs and the explanatory variables. Here, as in continuously updated GMM, $\beta_0$ appears both in the estimating function (through the constrained signs) and the weighting matrix.

Minimizing $D_S\big[\beta_0,\, \bar{J}_n(\beta_0)^{-1}\big]$ in (3.3) requires one to invert a new matrix $\bar{J}_n(\beta_0)$ for each value of $\beta_0$, whereas this is not needed for $D_S(\beta_0,\, I_p)$ or $D_S\big[\beta_0,\, (X'X)^{-1}\big]$. In practice, as for continuously updated GMM, this numerical problem may be cumbersome. To simplify calculations, it is also possible to use a two-step method: first, we solve (3.3) to obtain $\hat{\beta}_n = \hat{\beta}_n\big(y,\, X,\, SF\big)$; we then compute $\hat{J}_n\big[s(y - X\hat{\beta}_n),\, X\big]$ and minimize

$$D_S\big[\beta_0,\, \bar{J}_n(\hat{\beta}_n)^{-1}\big] = s(y - X\beta_0)'X\big[\hat{J}_n(s(y - X\hat{\beta}_n),\, X)\big]^{-1}X's(y - X\beta_0) \tag{3.6}$$

with respect to $\beta_0$. The estimator obtained in this way will be called hereafter the $SHAC$ sign-based estimator. Note however that no finite-sample distributional theory is available for $D_S\big[\beta_0,\, \bar{J}_n(\hat{\beta}_n)^{-1}\big]$, even under the mediangale assumption.

For heteroskedastic independent disturbances, we consider weighted versions of sign-based estimators which can be more efficient than the basic ones defined in (3.2) or (3.3). Weighted sign-based estimators are sign-based analogues to weighted LAD estimator [Zhao (2001)]. The weighted LAD estimator is given by

$$\beta_n^{WLAD} = \underset{\beta\in\mathbf{R}^p}{\operatorname{argmin}} \sum_i d_i|y_i - x_i'\beta|. \tag{3.7}$$

Correspondingly, we consider *scale weighted sign-based estimators* and *density weighted sign-based estimators*. A scale weighted sign-based estimator $\big[\hat{\beta}_n(H_n)\big]$ is obtained by minimizing

$$D_S\big[\beta_0,\, H_n\big] = s(y-X\beta_0)'XH_nX's(y-X\beta_0) = s(y-X\beta_0)'\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'s(y-X\beta_0) \tag{3.8}$$

where $H_n = D_n(\tilde{X}'\tilde{X})^{-1}D_n$, $\tilde{X} = XD_n$, and $D_n = \operatorname{diag}(d_1, \ldots, d_n)$ with $d_i > 0$, $i = 1, \ldots, n$. The density weighted sign-based estimator $\big[\hat{\beta}_n(H_n^*)\big]$ is based on optimal estimating functions [in the sense of Godambe (2001)] and minimizes

$$D_S\big[\beta_0,\, H_n^*\big] \quad = \quad s(y - X\beta_0)'XH_n^*X's(y - X\beta_0)$$

(1997, Section 3.1), Chen, Linton and Van Keilegom (2003), and Honore and Hu (2004).

$$= \quad s(y - X\beta_0)' X^* (X^{*\prime} X^*)^{-1} X^{*\prime} s(y - X\beta_0) \tag{3.9}$$

where $H_n^* = D_n^* (X^{*\prime} X^*)^{-1} D_n^*$, $X^* = X D_n^*$, $D_n^* = \mathrm{diag}\big[f_1(0|X), \dots, f_n(0|X)\big] X$, and $f_i(0|X)$ is the density of $u_t$ evaluated at zero (conditional on $X$), $i = 1, \dots, n$. An inherent difficulty for such estimators consists in approximating the density values $f_1(0|X), \dots, f_n(0|X)$. Note however that level can still be controlled, even if a conventional density (such as Gaussian density) is used .

Further, we show that under an additional weak mediangale assumption, the sign-based estimators presented here are equal (in probability) to Hodges-Lehmann estimators associated to the finite-sample sign-based testing theory developed in Coudin and Dufour (2009).

### 3.2. Hodges-Lehmann sign-based estimators

The estimators proposed above are closely related with the method proposed by Hodges and Lehmann (1963) to build point estimates from distribution-free tests on a scalar parameter; see also Johnson, Kotz and Read (1983). Suppose $\mu \in \mathbb{R}$ and $T(\mu_0, W)$ is a statistic for testing $\mu = \mu_0$ against $\mu > \mu_0$ based on the observations $W$. Suppose further that $T(\mu, W)$ is nondecreasing in the scalar $\mu$. Given a known central value of $T(\mu_0, W)$, say $m(\mu_0)$ [for example $\mathsf{E}_W T(\mu_0, W)$], the test rejects $\mu = \mu_0$ whenever the observed $T$ is larger than, say, $m(\mu_0)$. If this is the case, one is inclined to prefer higher values of $\mu$. The reverse holds when testing the opposite. If $m(\mu_0)$ does not depend on $\mu_0$ [$m(\mu_0) = m_0$], an intuitive estimator of $\mu$ (if it exists) is given by $\mu^*$ such that $T(\mu^*, W)$ equals $m_0$ (or is very close to $m_0$). $\mu^*$ may be seen as the value of $\mu$ which is most supported by the observations.

Here we consider an extension to multidimensional parameters through $p$-value functions. Let $\beta_0 \in \Theta$. Consider now testing $H_0(\beta_0) : \beta = \beta_0$ versus $H_1(\beta_0) : \beta \neq \beta_0$ using the test statistic $D_S[\beta_0, \bar{\Omega}_n(\beta_0)]$. A test based on $D_S$ rejects $H_0(\beta_0)$ when $D_S[\beta_0, \bar{\Omega}_n(\beta_0)]$ is larger than a certain critical value which depends on the test level. The estimator of $\beta$ is chosen as the value of $\beta$ least rejected when the level $\alpha$ of the test increases. This corresponds to the highest $p$-value. If the associated $p$-value for $H_0(\beta_0)$ is $\mathrm{p}(\beta_0) = G\big(D_S[\beta_0, \bar{\Omega}_n(\beta_0)]|\beta_0\big)$, where $G(x|\beta_0)$ is the survival function of $D_S[\beta_0, \bar{\Omega}_n(\beta_0)]$, *i.e.* $G(x|\beta_0) = \mathsf{P}[D_S[\beta_0, \bar{\Omega}_n(\beta_0)] > x \,|\, \beta = \beta_0]$, the set

$$M_1 = \underset{\beta_0 \in \Theta}{\arg\max}\; \mathrm{p}(\beta_0) \tag{3.10}$$

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

When the distribution of $D_S[\beta_0, \bar{\Omega}_n(\beta_0)]$ and the corresponding $p$-value function do not depend on the tested value $\beta_0$, maximizing the $p$-value is equivalent to minimizing the statistic $D_S[\beta_0, \bar{\Omega}_n(\beta_0)]$. This point is stated in the following proposition. Let us denote $\bar{F}(x|\beta_0)$ the distribution of $D_S[\beta_0, \bar{\Omega}_n(\beta_0)]$ when $\beta = \beta_0$ and assume this distribution is invariant to $\beta$ (Assumption 3.1).

**Assumption 3.1** INVARIANCE OF THE DISTRIBUTION FUNCTION.

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

Let us define

$$M_2 = \underset{\beta_0 \in \Theta}{\arg\min}\ D_S(\beta_0, \Omega_n). \tag{3.11}$$

Then, the following proposition holds.

**Proposition 3.1** *If Assumption 3.1 holds, then $M_1 = M_2$ with probability one.*

If the disturbances satisfy the mediangale Assumption 2.1, any sign-based statistic constitutes a pivotal function under $H_0(\beta_0)$; see Coudin and Dufour (2009). Hence, Assumption 3.1 is satisfied and $\hat{\beta}_n(\Omega_n)$ can be viewed as a Hodges-Lehmann estimator based on $D_S(\Omega_n, \beta)$.

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

Sign-based estimators have usually been interpreted in the literature as GMM estimators exploiting the orthogonality condition between the signs and the explanatory variables or instruments [see Honore and Hu (2004)]. However, the GMM interpretation hides the link with testing theory, which is revealed by the Hodges-Lehmann estimator interpretation. Hodges-Lehmann estimators correspond to parameter values which are least rejected by the tests (given the data). 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. The finite-sample properties of sign-based estimators are studied in the next section.

## 4. Finite-sample properties of sign-based estimators

In this section, finite-sample properties of sign-based estimators are studied. Sign-based estimators share invariance properties with the LAD estimator and are median-unbiased if the disturbance distribution is symmetric and some additional assumptions on the form of the solution are satisfied.

The topology of the argmin set of the optimization problem 3.1 does not possess a simple structure. In some cases it is reduced to a single point like the empirical median of $2p+1$ observations. In other cases, it is a set. More generally, the argmin set is a union of convex sets but it is not *a priori* either convex nor connected. To see that it is a union of convex sets just remark that the reciprocal image of $n$ fixed signs is convex.

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

**Proposition 4.1** INVARIANCE. *Let $M(y, X)$ be the set of the solutions of the minimization problem $(3.1)$. If $\hat{\beta}(y, X) \in M(y, X)$, then the following properties hold:*

$$\lambda\hat{\beta}(y, X) \in M(\lambda y, X), \quad \forall\lambda \in \mathbb{R}, \tag{4.1}$$

$$\hat{\beta}(y, X) + \gamma \in M(y + X\gamma, X), \quad \forall\gamma \in \mathbb{R}^p, \tag{4.2}$$

$$A^{-1}\hat{\beta}(y, X) \in M(y, XA), \quad \textit{for any nonsingular } k \times k \textit{ matrix } A. \tag{4.3}$$

*Further, if $\hat{\beta}(y, X)$ is a uniquely determined solution of $(3.1)$, then*

$$\hat{\beta}(\lambda y, X) = \lambda\hat{\beta}(y, X), \quad \forall\lambda \in \mathbb{R}, \tag{4.4}$$

$$\hat{\beta}(y + X\gamma, X) = \hat{\beta}(y, X) + \gamma, \quad \forall\gamma \in \mathbb{R}^p, \tag{4.5}$$

$$\hat{\beta}(y, XA) = A^{-1}\hat{\beta}(y, X), \quad \textit{for any nonsingular } k \times k \textit{ matrix } A. \tag{4.6}$$

To prove this property, it is sufficient to write down the different optimization problems. (4.1) and (4.4) 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. (4.2) and (4.5) represent location invariance, while (4.3) and (4.6) show the behavior of the estimator changes states a reparameterization of the design matrix. In all cases, parameter estimates change in the same way as theoretical parameters.

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

**Proposition 4.2** MEDIAN UNBIASEDNESS. *If $u \sim -u$ and the sign-based estimator $\hat{\beta}(y, X)$ is a uniquely determined solution of the minimization problem$(3.1)$, then $\hat{\beta}$ is median unbiased,* i.e.

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

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

# 5. Asymptotic properties

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

## 5.1. Identification and consistency

We show that the sign-based estimators (3.1) and (3.6) are consistent under the following set of assumptions. In the sequel, we denote by $\bar{\beta}$ the "true value" of $\beta$, and by $\beta_0$ any hypothesized value.

**Assumption 5.1** MIXING. *$\{W_t = (y_t, x_t')\}_{t=1,2,\ldots}$ is $\alpha$-mixing of size $-r/(r-1)$ with $r > 1$.*

**Assumption 5.2** 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 5.3** COMPACTNESS. *$\bar{\beta} \in \mathrm{Int}(\Theta)$, where $\Theta$ is a compact subset of $\mathbb{R}^p$.*

**Assumption 5.4** REGULARITY OF THE DENSITY.

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

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

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

**Assumption 5.5** POINT IDENTIFICATION CONDITION. *$\forall \delta > 0, \exists \tau > 0$ such that*

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

**Assumption 5.6** UNIFORMLY POSITIVE DEFINITE WEIGHT MATRIX. *$\bar{\Omega}_n(\beta)$ is symmetric positive definite for all $\beta$ in $\Theta$.*

**Assumption 5.7** LOCALLY POSITIVE DEFINITE WEIGHT MATRIX. *$\bar{\Omega}_n(\beta)$ is symmetric positive definite for all $\beta$ in a neighborhood of $\bar{\beta}$.*

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

**Theorem 5.1** CONSISTENCY. *Under model* $(2.1)$ *with the assumptions 2.2 and 5.1-5.6, any sign-based estimator of the type,*

$$\hat{\beta}_n \in \underset{\beta_0 \in \Theta}{\operatorname{argmin}}\, s(y - X\beta_0)' X \Omega_n \big[s(y - X\beta_0), X\big] X' s(y - X\beta_0) \tag{5.1}$$

*or*

$$\hat{\beta}_n^{2S} \in \underset{\beta_0 \in \Theta}{\operatorname{argmin}}\, s(y - X\beta_0)' X \hat{\Omega}_n \big[s(y - X\hat{\beta}), X\big] X' s(y - X\beta_0), \tag{5.2}$$

*where* $\hat{\beta}$ *stands for any (first step) consistent estimator of* $\bar{\beta}$*, is consistent.* $\hat{\beta}_n^{2S}$ *defined in equation* $(5.2)$ *is also consistent if Assumption 5.6 is replaced by Assumption 5.7.*

It will useful to discuss Assumptions 5.1 - 5.7 and compare them to the ones required for LAD and quantile estimator consistency; see Fitzenberger (1997) and Weiss (1991). The mixing assumption 5.1 is needed to apply a generic weak law of large numbers; see Andrews (1987) and White (2001). It was used by Fitzenberger (1997) to show LAD and quantile estimator consistency with stationary linearly dependent processes. It covers, among other processes, stationary ARMA disturbances with continuously distributed innovations. Point identification is provided by assumptions 5.4 and 5.5. Assumption 5.5 is similar to Condition ID in Weiss (1991). Assumption 5.4 is usual in LAD estimator asymptotics.[6] It is analogous to Fitzenberger's (1997) conditions (ii.b) - (ii.c) and Weiss's (1991) CD condition. It implies that there is enough variation around zero to identify the median. It restricts the setup for some "bounded" heteroskedasticity in the disturbance process but not in the usual (variance-based) way. It is related to *diffusivity* $\frac{1}{2f(0)}$ , an alternative measure of dispersion adapted to median-unbiased estimators. Diffusivity measures the vertical spread of a density rather than its horizontal spread, and appears in Cramér-Rao-type lower bound for median-unbiased estimators; see Sung, Stangenhaus and David (1990) and So (1994). Assumption 5.6 entails that the weight matrix $\Omega_n$ is everywhere invertible, while Assumption 5.7 only requires local invertibility.

An important difference with the LAD asymptotic theory comes from Assumption 5.2. For sign consistency, only the second-order moments of $x_t$ have to be finite, which differs from Fitzenberger (1997) who assumed the existence of at least third-order moments. We do not assume the existence of second-order moments on the disturbances $u_t$. The disturbances indeed appear in the objective function only through their sign transforms which possess finite moments up to any order. Consequently, no additional restriction should be imposed on the disturbance process (in addition to regularity conditions on the density). Those points will entail a more general CLT than the one

[6] Assumption 5.4 can be slightly relaxed covering error terms with mass point if the objective function involves randomized signs instead of usual signs.

stated for the LAD/quantile estimators in Fitzenberger (1997) and Weiss (1991). The only works we are aware of that study LAD estimators properties in case of infinite variance errors are those of 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.

### 5.2. Asymptotic normality

Sign-based estimators are asymptotically normal. This also holds under weaker assumptions than the ones needed for LAD estimator asymptotic normality as presented in Weiss (1991) and Fitzenberger (1997). Sign-based estimators are well adapted to deal with heavy-tailed disturbances that may not possess finite variances. The assumptions we consider are the following ones.

**Assumption 5.8** UNIFORMLY BOUNDED DENSITIES. $\exists f_U < +\infty$ *such that* $, \forall n \in \mathbb{N}, \forall \lambda \in \mathbb{R}$*,*

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

Under the conditions 2.2, 5.1, 5.2 and 5.8, we can define $L(\beta)$, the derivative of the limiting objective function at $\beta$:

$$L(\beta) = \lim_{n\to\infty} \frac{1}{n} \sum_t \mathsf{E}\big[x_t x_t' f_t\big(x_t'(\beta - \bar{\beta}) \,|\, x_1, \ldots, x_n\big)\big] = \lim_{n\to\infty} L_n(\beta). \tag{5.3}$$

where

$$L_n(\beta) = \frac{1}{n} \sum_t \mathsf{E}\big[x_t x_t' f_t\big(x_t'(\beta - \bar{\beta}) \,|\, x_1, \ldots, x_n\big)\big]. \tag{5.4}$$

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

**Assumption 5.9** MIXING WITH $r > 2$. *The process* $\{W_t = (y_t, x_t') : t = 1, 2, \ldots\}$ *is* $\alpha$*-mixing of size* $-r/(r-2)$ *with* $r > 2$*.*

**Assumption 5.10** DEFINITE POSITIVENESS OF $L_n$. $L_n(\bar{\beta})$ *is positive definite uniformly in* $n$*.*

**Assumption 5.11** DEFINITE POSITIVENESS OF $J_n$. $J_n = \mathsf{E}\big[\frac{1}{n}\sum_{t,s}^{n} s(u_t) x_t x_s' s(u_s)\big]$ *is positive definite uniformly in* $n$ *and converges to a definite positive symmetric matrix* $J$ *as* $n \to \infty$*.*

Then, we have the following result.

**Theorem 5.2** ASYMPTOTIC NORMALITY. *Under the assumptions* (2.2)*, 5.1 to 5.6, and 5.9 to 5.11, we have:*

$$S_n^{-1/2} \sqrt{n} \big[\hat{\beta}_n - \bar{\beta}\big] \xrightarrow{d} \mathrm{N}(0, I_p) \tag{5.5}$$

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

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

*and*

$$L_n(\bar{\beta}) = \frac{1}{n}\sum_t \mathsf{E}\big[x_t x_t' f_t\big(0 \,|\, x_1, \ldots, x_n\big)\big]. \tag{5.6}$$

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

$$[L_n(\bar{\beta})\hat{J}_n^{-1}L_n(\bar{\beta})]^{-1/2}\sqrt{n}\big[\hat{\beta}_n(\hat{J}_n^{-1}) - \bar{\beta}\big] \xrightarrow{d} \mathrm{N}\big[0, I_p\big]\,. \tag{5.7}$$

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

Note again that we do not require that the disturbance process variance be finite. We only assume that the second-order moments of $X$ are finite and the mixing property of $\{W_t,\ t = 1, \ldots\}$ holds. This differs from usual assumptions for LAD asymptotic normality.[7] This difference comes from the fact that absolute values of the disturbance process are replaced in the objective function by their signs. Since signs possess finite moments at any order, one sees easily that a CLT can be applied without any further restriction. Consequently, asymptotic normality, such as consistency, holds for heavy-tailed disturbances that may not possess finite variance. This is an important theoretical advantage of sign-based rather than absolute value-based estimators and, *a fortiori*, rather than least-squares estimators. Estimators, for which asymptotic normality holds on bounded asymptotic variance assumption (for example OLS) are not accurate in heavy-tail settings because the variance is not a measure of dispersion adapted to those settings. Estimators, for which the asymptotic behavior relies on other measures of dispersion, like the diffusivity, help one out of trouble.

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

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

[7] 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.

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

$$S_n = \frac{1}{4f(0)^2}\mathsf{E}(x_t x_t')^{-1}. \tag{5.8}$$

In the general case, $f_t(0)$ is a nuisance parameter even if condition 5.8 implies that it can be bounded.

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

### 5.3. Asymptotic or projection-based confidence sets?

In section 3, we introduced sign-based estimators as Hodges-Lehmann estimators associated with sign-based statistics. By linking them with GMM settings, we then derived asymptotic normality. We stressed that sign-based estimator asymptotic normality holds under weaker assumptions than the ones needed for the LAD estimator. Therefore, sign-based estimator asymptotic normality enables one to construct asymptotic tests and confidence intervals. Thus, we have two ways of making inference with signs: we can use the Monte Carlo (finite-sample) based method described in Coudin and Dufour (2009) - see section 2.2 - and the classical asymptotic method. Let us list here the main differences between them. Monte Carlo inference relies on the pivotality of the sign-based statistic. The derived tests are valid (with controlled level) for any sample size if the mediangale Assumption 2.1 holds. When only the sign moment condition 2.2 holds, the Monte Carlo inference remains asymptotically valid. Asymptotic test levels are controlled. Besides, in simulations, the Monte Carlo inference method appears to perform better in small samples than classical asymptotic methods, even if its use is only asymptotically justified [see Coudin and Dufour (2009)]. Nevertheless, that method has an important drawback: its computational complexity. On the contrary, classical asymptotic methods which yield tests with controlled asymptotic level under the sign moment condition 2.2 may be less time consuming. The choice between both is mainly a question of computational capacity. We point out that classical asymptotic inference greatly relies on the way the asymptotic covariance matrix, that depends on unknown parameters (densities at zero), is treated. If the asymptotic covariance matrix is estimated thanks to a simulation-based method (such as the bootstrap) then the time argument does not hold anymore. Both methods would be of the same order of computational complexity.

# 6. Simulation study

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

## 6.1. Simulation setup

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

$$y_t = x_t'\beta + u_t \tag{6.1}$$

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

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

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

## 6.2. Bias and RMSE

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

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

Table 1. Simulated models.

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

Table 2. Simulated bias and RMSE.

| $n = 50$ | | OLS | | LAD | | SF | | SHAC | |
|---|---|---|---|---|---|---|---|---|---|
| $S = 1000$ | | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE |
| A1: | $\beta_0$ | .003 | .142 | .002 | .179 | .002 | .179 | .004 | .178 |
| | $\beta_1$ | .003 | .149 | .006 | .184 | .004 | .182 | .004 | .182 |
| | $\beta_2$ | $-.002$ | .149 | $-.007$ | .186 | $-.006$ | .185 | $-.007$ | .183 |
| | $\|\|\boldsymbol{\beta}\|\|$* | $\mathbf{.004}$ | $\mathbf{.254}$ | $\mathbf{.009}$ | $\mathbf{.316}$ | $\mathbf{.007}$ | $\mathbf{.315}$ | $\mathbf{.009}$ | $\mathbf{.313}$ |
| A2: | $\beta_0$ | $-.003$ | .136 | .000 | .090 | $-.000$ | .089 | $-.000$ | .089 |
| | $\beta_1$ | $-.0135$ | .230 | $-.006$ | .218 | $-.010$ | .218 | $-.010$ | .218 |
| | $\beta_2$ | .002 | .142 | $-.001$ | .095 | $-.001$ | .092 | $-.001$ | .092 |
| | $\|\|\boldsymbol{\beta}\|\|$ | $\mathbf{.014}$ | $\mathbf{.303}$ | $\mathbf{.007}$ | $\mathbf{.254}$ | $\mathbf{.010}$ | $\mathbf{.253}$ | $\mathbf{.010}$ | $\mathbf{.253}$ |
| A3: | $\beta_0$ | .022 | .167 | .018 | .108 | .025 | .107 | .023 | .107 |
| | $\beta_1$ | $-1.00$ | .228 | .005 | .215 | .003 | .214 | .002 | .215 |
| | $\beta_2$ | .001 | .150 | .005 | .105 | .007 | .104 | .007 | .105 |
| | $\|\|\boldsymbol{\beta}\|\|$ | $\mathbf{.022}$ | $\mathbf{.320}$ | $\mathbf{.019}$ | $\mathbf{.263}$ | $\mathbf{.026}$ | $\mathbf{.261}$ | $\mathbf{.024}$ | $\mathbf{.262}$ |
| A4: | $\beta_0$ | $-.001$ | .174 | .007 | .2102 | .010 | .2181 | .008 | .2171 |
| | $\beta_1$ | $-.016$ | .313 | $-.011$ | .375 | $-.021$ | .396 | $-.021$ | .394 |
| | $\beta_2$ | $-.100$ | 14.6 | .077 | 18.4 | .014 | 7.41 | .049 | 7.40 |
| | $\|\|\boldsymbol{\beta}\|\|$ | $\mathbf{.101}$ | $\mathbf{14.6}$ | $\mathbf{.078}$ | $\mathbf{18.5}$ | $\mathbf{.027}$ | $\mathbf{7.42}$ | $\mathbf{.054}$ | $\mathbf{7.41}$ |
| B5: | $\beta_0$ | 16.0 | 505 | .001 | .251 | .004 | .248 | .003 | .248 |
| | $\beta_1$ | $-3.31$ | 119 | .015 | .264 | .020 | .265 | .020 | .265 |
| | $\beta_2$ | $-2.191$ | 630 | .000 | .256 | .003 | .258 | .001 | .258 |
| | $\|\|\boldsymbol{\beta}\|\|$ | $\mathbf{26.0}$ | $\mathbf{817}$ | $\mathbf{.015}$ | $\mathbf{.445}$ | $\mathbf{.021}$ | $\mathbf{.445}$ | $\mathbf{.020}$ | $\mathbf{.445}$ |
| B6: | $\beta_0$ | $-.908$ | 29.6 | $-1.02$ | 27.4 | .071 | 2.28 | .083 | 2.28 |
| | $\beta_1$ | 2.00 | 37.6 | 3.21 | 68.4 | .058 | 2.38 | .069 | 2.39 |
| | $\beta_2$ | 1.64 | 59.3 | 2.59 | 91.8 | $-.101$ | 2.30 | $-.089$ | 2.29 |
| | $\|\|\boldsymbol{\beta}\|\|$ | $\mathbf{2.73}$ | $\mathbf{76.2}$ | $\mathbf{4.25}$ | $\mathbf{118}$ | $\mathbf{.136}$ | $\mathbf{4.02}$ | $\mathbf{.139}$ | $\mathbf{4.02}$ |
| B7: | $\beta_0$ | $-127$ | 3289 | $-.010$ | 7.85 | $-.008$ | 3.16 | $-.028$ | 3.17 |
| | $\beta_1$ | $-81.4$ | 237 | .130 | 11.2 | $-.086$ | 3.80 | $-.086$ | 3.823 |
| | $\beta_2$ | $-31.0$ | 1484 | $-.314$ | 12.0 | $-.021$ | 3.606 | $-.009$ | 3.630 |
| | $\|\|\boldsymbol{\beta}\|\|$ | $\mathbf{154}$ | $\mathbf{4312}$ | $\mathbf{.340}$ | $\mathbf{18.2}$ | $\mathbf{.089}$ | $\mathbf{6.12}$ | $\mathbf{.091}$ | $\mathbf{6.15}$ |
| B8: | $\beta_0$ | $< -10^{10}$ | $> 10^{10}$ | $< -10^{9}$ | $> 10^{10}$ | .312 | 5.67 | .307 | 5.67 |
| | $\beta_1$ | $> 10^{10}$ | $> 10^{10}$ | $> 10^{9}$ | $> 10^{10}$ | .782 | 5.40 | .863 | 5.46 |
| | $\beta_2$ | $< -10^{10}$ | $> 10^{10}$ | $< -10^{9}$ | $> 10^{10}$ | .696 | 5.52 | .696 | 5.55 |
| | $\|\|\boldsymbol{\beta}\|\|$ | $> \mathbf{10^{10}}$ | $> \mathbf{10^{10}}$ | $> \mathbf{10^{10}}$ | $> \mathbf{10^{10}}$ | $\mathbf{1.09}$ | $\mathbf{9.58}$ | $\mathbf{1.15}$ | $\mathbf{9.63}$ |
| C9: | $\beta_0$ | .005 | .279 | .001 | .308 | .003 | .309 | .004 | .311 |
| | $\beta_1$ | $-.002$ | .163 | $-.005$ | .201 | $-.004$ | .200 | $-.005$ | .199 |
| | $\beta_2$ | .001 | .165 | $-.004$ | .204 | .003 | .198 | .002 | .198 |
| | $\|\|\boldsymbol{\beta}\|\|$ | $\mathbf{.006}$ | $\mathbf{.363}$ | $\mathbf{.007}$ | $\mathbf{.420}$ | $\mathbf{.006}$ | $\mathbf{.418}$ | $\mathbf{.006}$ | $\mathbf{.419}$ |
| C10: | $\beta_0$ | $-.013$ | .284 | $-.010$ | .315 | $-.015$ | .314 | $-.014$ | .314 |
| | $\beta_1$ | $-.009$ | .182 | $-.009$ | .220 | $-.011$ | .218 | $-.011$ | .219 |
| | $\beta_2$ | .008 | .189 | .011 | .222 | .007 | .215 | .007 | .215 |
| | $\|\|\boldsymbol{\beta}\|\|$ | $\mathbf{.018}$ | $\mathbf{.387}$ | $\mathbf{.018}$ | $\mathbf{.444}$ | $\mathbf{.020}$ | $\mathbf{.439}$ | $\mathbf{.019}$ | $\mathbf{.439}$ |
| C11: | $\beta_0$ | .070 | 1.23 | $-.026$ | .308 | .058 | 1.26 | .053 | 1.27 |
| | $\beta_1$ | $-.000$ | .268 | .005 | .214 | $-.005$ | .351 | $-.008$ | .354 |
| | $\beta_2$ | .001 | .273 | $-.004$ | .210 | .002 | .361 | $-.001$ | .361 |
| | $\|\|\boldsymbol{\beta}\|\|$ | $\mathbf{.070}$ | $\mathbf{1.29}$ | $\mathbf{.027}$ | $\mathbf{.430}$ | $\mathbf{.059}$ | $\mathbf{1.36}$ | $\mathbf{.054}$ | $\mathbf{1.37}$ |

* ||.|| stands for the Euclidean norm.

estimator is notably smaller than those of the OLS and the LAD estimates.

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

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

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

# 7. Empirical application: drift estimation with stochastic volatility in the error term

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

$$y_t = a + bt + u_t, \quad t = 1, \dots, 16127, \tag{7.1}$$

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

We compute both the basic $SF$ sign-based estimator and the $SHAC$ version with the two-step

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

Table 3. Constant and drift estimates.

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

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

** 95% confidence intervals.

method. They are compared with the LAD and OLS estimates. Then, we redo a similar experiment on two subperiods: on the year 1929 (291 observations) and the last 90 days of 1929, which roughly corresponds to the four last months of 1929 (90 observations). Due to the financial crisis, one may expect data to involve an extreme amount of heteroskedasticity in that period of time. We wonder at which point that heteroskedasticity can bias the subsample estimates. The Wall Street crash occurred between October, 24th (*Black Thursday*) and October, 29th (*Black Tuesday*). Hence, the second subsample corresponds to the period just before the krach (September), the krach period (October) and the early beginning of the Great Depression (November and December). Heteroskedasticity tests reject homoskedasticity for both subsamples.[9]

In Table 3, we report estimates and recall the $95\%$ confidence intervals for $a$ and $b$ obtained by the finite-sample sign-based method ($SF$ and $SHAC$);[10] and by moving block bootstrap (LAD and OLS). The entire set of sign-based estimators is reported, *i.e.*, all the minimizers of the sign objective function.

[9] 1929: White: 24.2, $p$-values: .000 ; BP: 126, $p$-values: .000; Sept-Oct-Nov-Dec 1929: White: 11.08, $p$-values: .004; BP: 1.76, $p$-values: .18.

[10] see Coudin and Dufour (2009)

First, we note that the OLS estimates are importantly biased and are greatly unreliable in the presence of heteroskedasticity. Hence, they are just reported for comparison sake. Presenting the entire sets of sign-based estimators enables us to compare them with the LAD estimator. In this example, LAD and sign-based estimators yield very similar estimates. The value of the LAD estimator is indeed just at the limit of the sets of sign-based estimators. This does not mean that the LAD estimator is included in the set of sign-based estimators, but, there is a sign-based estimator giving the same value as the LAD estimate for a certain individual component (the second component may differs). One easy way to check this is to compare the two objective functions evaluated at the two estimates. For example, in the 90 observation sample, the sign objective function evaluated at the basic sign-estimators is $4.75 \times 10^{-3}$, and at the LAD estimate $5.10 \times 10^{-2}$; the LAD objective function evaluated at the LAD estimate is $210.4$ and at one of the sign-based estimates $210.5$. Both are close but different.

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

## 8. Conclusion

In this paper, we have introduced a class of robust sign-based estimators for the parameters of a linear median regression. We have shown that they turn out to be equivalent (in probability) to Hodges-Lehmann estimators when a mediangale assumption holds. In such a case they are the parameter values the less rejected by finite-sample distribution-free sign-based tests. Hence, they are derived without referring to asymptotic conditions through the analogy principle. Then we have presented general properties of sign-based estimators (invariance, median unbiasedness) and the conditions under which consistency and asymptotic normality hold. In particular, we have shown that sign-based estimators do require less assumptions on moment existence of the disturbances than usual LAD asymptotic theory. Simulation studies indicate that the proposed estimators are accurate in classical setups and more reliable than usual methods (LS, LAD) when arbitrary heterogeneity or nonlinear dependence is present in the error term even in cases that may cause LAD or OLS consistency failure. Despite the programming complexity of sign-based methods, we recommend combining sign-based estimators to the Monte Carlo sign-based method of inference presented in Coudin and Dufour (2009) when an amount of heteroskedasticity is suspected in the data and when the number of available observations is small. As illustrative application, we estimate a drift parameter on the Standard and Poor's Composite Price Index, using the 1928-1987 period and various shorter subperiods.

# Appendix

## A. Proofs

**Proof of Proposition 3.1**. Let $D_S$ be a sign-based statistic of the form presented in equation (3.1). The term $\Omega_n$ is omitted for simplicity. We show that the sets $M_1$ and $M_2$ are equal with probability one. First, we show that if $\hat{\beta} \in M_2$ then it belongs to $M_1$. Second, we show that if $\hat{\beta}$ does not belong to $M_2$, neither it belongs to $M_1$.
If $\hat{\beta} \in M_2$ then,

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

hence

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

and $\hat{\beta}$ maximizes the $p$-value. Conversely, if $\hat{\beta}$ does not belong to $M_1$, there is a non negligible Borel set, say $A$, such that $D_S(\beta) < D_S(\hat{\beta})$ on $A$ for some $\beta$. 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 $\beta$ (Assumption 3.1),

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

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 4.2**. Consider $\hat{\beta}(y, X, u)$ the solution of problem (3.1) which is assumed to be unique, let $\bar{\beta}$ be the true value of the parameter $\beta$ and suppose that $u \sim -u$. Equation (4.4) implies that

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

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

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

Moreover, equation (4.5) implies that

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

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

**Proof of Theorem 5.1**. 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 $(\Omega,\mathcal{F},\mathsf{P})$. We denote

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

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

**Pointwise convergence**. The mixing property 5.1 on $W$ is exported to $\{q_{tk}(W_t,\beta),\ \ k=1,\ldots,p\}_{t=1,2,\ldots}$. Hence, $\forall\beta\in\Theta,\ \ \forall k=1,\ldots,p,\ \ \{q_{tk}(W_t,\beta)\}$ is an $\alpha-$mixing process of size $r/(1-r)$. Moreover, condition 5.2 entails $\mathsf{E}|q_{tk}(W_t,\beta)|^{r+\delta}<\infty$ for some $\delta>0$, for all $t\in\mathbb{N}$, $k=1,\ldots,p$. Hence, we can apply Corollary 3.48 of White (2001) to $\{q_{tk}(W_t,\beta)\}_{t=1,2,\ldots}$. It follows $\forall\beta\in\Theta$,

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

**Uniform Convergence**. We check conditions A1, A6, B1, B2 of Andrews (1987)'s generic weak law of large numbers (GWLLN). A1 and B1 are our conditions 5.3 and 5.1. Then, Andrews defines

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

where $B(\beta,\rho)$ is the open ball around $\beta$ of radius $\rho$. His condition B2 requires that $q^H_{tk}(W_t,\beta,\rho)$, $q_{Ltk}(W_t,\beta,\rho)$ and $q_{tk}(W_t)$ are random variables; $q^H_{tk}(.,\beta,\rho)$, $q_{Ltk}(.,\beta,\rho)$ are measurable functions from $(\Omega,\mathcal{P},\mathcal{F})$ to $(\mathbb{R},\mathcal{B})$, $\forall t,\ \ \beta\in\Theta,\ \ \rho$, where $\mathcal{B}$ is the Borel $\sigma$-algebra on $\mathbb{R}$ and finally, that $\sup_t \mathsf{E}q_{tk}(W_t)^{\xi}<\infty$ with $\xi>r$. Those points are derived from the mixing condition 5.1 and condition 5.2 which insures measurability and provides bounded arguments.

The last condition (A6) to check requires the following: Let $\mu$ be a $\sigma$-finite measure that dominates each one of the marginal distributions of $W_t,\ \ t=1,2\ldots$. Let $p_t(w)$ be the density of $W_t$ w.r.t. $\mu$, $q_{tk}(W_t,\ \beta)p_t(W_t)$ is continuous in $\beta$ at $\beta=\beta^*$ uniformly in $t$ a.e. w.r.t. $\mu$, for each

$\beta^* \in \Theta$, $q_{tk}(W_t,\ \beta)$ is measurable w.r.t. the Borel measure for each $t$ and each $\beta \in \Theta$, and $\int \sup_{t\geq 0,\ \beta\in\Theta} |q_{tk}(W,\ \beta)| p_t(w) d\mu(w) < \infty$. As $u_t$ is continuously distributed uniformly in $t$ [Assumption 5.4 (2)], we have $\mathsf{P}_t[u_t = x_t\beta] = 0,\ \forall\beta$, uniformly in $t$. Then, $q_{tk}$ is continuous in $\beta$ everywhere except on a $\mathsf{P}_t$-negligible set. Finally, since $q_{tk}$ is $L_1$-bounded and uniformly integrable, condition A6 holds.

The generic law of large numbers (GWLLN) implies:

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

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

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

$$\lim_{n\to\infty} \frac{1}{n}\sum_{t} \mathsf{E}[s(u_t - x_t'\beta)x_{tk}] \text{ is continuous on } \Theta, k = 1, \ldots, p. \tag{A.7}$$

Let us define

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

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

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

$$\frac{1}{n^2}\sum_{k=1}^{p}\left(\sum_{t} x_{kt} s(u_t - x_t'\beta)\right)^2 = \sum_{k} [Q_n^k(\beta)]^2.$$

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

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

Uniform weak convergence of $Q_n^k$ to $Q_n^{Ek}$ at $\beta_n$ implies:

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

hence,

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

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

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

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

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

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

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

Provided that 0 is the unique minimizer, we have:

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

Hence, setting

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

it follows that, with probability close to one,

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

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

For **identification**, the uniqueness of the minimizer of the sign-objective function is insured by the set of identification conditions 2.2, 5.5, 5.4, 5.6. 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 $\beta$ (condition 5.6), this assertion is equivalent to

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

and

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

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

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

Note that

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

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

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

Then,

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

$$+I_{\{|x_t'\delta|\le\tau\}}(I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta}-f_t(u|x_1,\ldots,x_n)du$$
$$+\ I_{\{x_t'\delta\le 0\}}\int_{x_t'\delta}^0 f_t(u|x_1,\ldots,x_n)du)\Bigg]\Bigg\}.$$

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

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

To obtain inequation (A.15), just remark that each term is positive. For the inequation (A.16) we use condition 5.4. For inequation (A.17) we minorate $|x_i'\delta|$ by $\tau$ and each integrals by $f_L d_1$ where $d_1=\min(\tau,d/2)$. Condition 5.5 enables us to conclude, by taking the limit,

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

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

**Proof of Theorem 5.2**. We prove Theorem **5.2** on asymptotic normality. We consider the sign-based estimator $\hat{\beta}(\Omega_n)$ where $\Omega_n$ 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 (5.3) is the derivative of $\lim_{n\to\infty} \frac{1}{n}\sum_t \mathsf{E}\big[s\big(u_t - x_t'(\beta - \bar{\beta})\big)x_t\big]$. Then, we check the conditions for applying Theorem 7.2 of Newey-McFadden.

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

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

is continuous on $\Theta$ uniformly over $n$. Moreover condition 5.2 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 $\lambda$ (condition 5.8), dominated convergence allows us to write that

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

And, these conditions imply that

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

converges uniformly in $\beta$ to $L(\beta)$. Uniform convergence entails that $\lim_n \frac{1}{n}\sum_{t=0}^{n} \mathsf{E}\big[s\big(u_t - x_t'(\beta - \bar{\beta})\big)x_t\big]$ is differentiable with derivative $L(\beta)$.
We now apply Theorem 7.2 of Newey and McFadden (1994) which presents asymptotic normality of a minimum distance consistent estimator with nonsmooth objective function and weight matrix $\Omega_n \xrightarrow{p} \Omega$ symmetric positive definite. Thus, under conditions for consistency (2.2, 5.1-5.6), we have to check that the following conditions hold:

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

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

$L(\bar{\beta})\Omega L(\bar{\beta})'$ is nonsingular;

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

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

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

Condition (i) is fulfilled by the moment condition 2.2. Condition (ii) is fulfilled by the first part of our proof and condition 5.10. Then, Condition (iii) is implied by 5.3. Using the mixing specification 5.9 of $\{u_t, X_t\}_{t=1,2,\dots}$ and conditions 2.2, 5.2, 5.7 and 5.11, we apply a White-Domowitz central limit theorem [see White (2001), Theorem 5.20]. This fulfills condition (iv) of Theorem 7.2 in Newey and McFadden (1994):

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

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

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

The asymptotic covariance matrix $S$ is given by the limit of

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

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

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

with

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

When the mediangale Assumption (2.1) holds, we find usual results on sign-based estimators. $\hat{\beta}(I_p)$ and $\hat{\beta}((X'X)^{-1})$ are asymptotically normal with asymptotic covariance matrix

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

□

## References


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

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

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

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

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

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

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

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

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

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

Chen, X., Linton, O. and Van Keilegom, I. (2003), 'Estimation of semiparametric models when the criterion function is not smooth', *Econometrica* **71**(5), 1591–1608.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Goffe, W. L., Ferrier, G. D. and Rogers, J. (1994), 'Global optimization of statistical functions with simulated annealing', *Journal of Econometrics* **60**(1), 65–100.

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

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

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

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

Honoré, B. E. and Powell, J. L. (1994), 'Pairwise difference estimators of censored and truncated regression models', *Journal of Econometrics* **64**, 241–278.

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.

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

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

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

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

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

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

Press, W., Teukolsky, S., Vetterling, W. and Flannery, B. (2002), *Numerical Recipes in Fortran 90*, Press Syndicate of the University of Cambridge, USA.

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

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

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

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

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

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