Finite and large sample distribution-free inference in linear median regressions under heteroskedasticity and nonlinear dependence of unknown form

Elise Coudin [*]
CREST and Université de Montréal

Jean-Marie Dufour [†]
Université de Montréal

First version: April 2003
Revised: August 2004
This version: May 2005
Compiled: May 27, 2005, 1:22pm

[*] Centre de Recherche en Economie et Statistique (CREST), and Département de sciences économiques, Université de Montréal. email: elise.coudin@ensae.fr

[†] Canada Research Chair Holder (Econometrics). Centre interuniversitaire de recherche en analyse des organisations (CIRANO), Centre interuniversitaire de recherche en économie quantitative (CIREQ), and Département de sciences économiques, Université de Montréal. Mailing address: Département de sciences économiques, Université de Montréal, C.P. 6128 succursale Centre-ville, Montréal, Québec, Canada H3C 3J7. TEL: 1 514 343 2400; FAX: 1 514 343 5831; e-mail: jean.marie.dufour@umontreal.ca. Web page: http://www.fas.umontreal.ca/SCECO/Dufour .

ABSTRACT

We study the construction of finite-sample distribution-free tests and the corresponding confidence sets for the parameters of a linear median regression where no parametric assumptions are imposed on the noise distribution. The setup we consider allows for non-normality, heteroskedasticity and nonlinear serial dependence of unknown forms, including non-Gaussian GARCH and stochastic volatility models with an unspecified order. Such semiparametric models are usually analyzed using only asymptotically justified approximate methods, which can be arbitrarily unreliable. We point out that statistics dealing with signs of residuals enjoy the required pivotality features – in addition to usual robustness properties. Then, sign-based statistics are exploited – in association with Monte-Carlo tests and projection techniques – in order to produce valid inference in finite samples. An asymptotic theory which holds under even weaker assumptions is also provided. Finally, simulation results illustrating the performance of the proposed methods are presented.

Key words: linear regression; non-normality; heteroskedasticity; serial dependence; GARCH; stochastic volatility; sign test; simultaneous inference; Monte Carlo tests; bootstrap; projection methods; quantile regressions.

Contents

List of Propositions and Theorems

1. Introduction

The Laplace-Boscovich median regression has received a renewed interest since two decades. This method is known to be more robust than the Least Squares method and specially adapted for heterogeneous data [see Dodge (1997)]. It has recently been adapted to models involving heteroskedasticity and autocorrelation [Zhao (2001), Weiss (1990)], endogeneity [Amemiya (1982), Powell (1983), Hong and Tamer (2003)], nonlinear functional forms [Weiss (1991)] and has been generalized to other quantile regressions [Koenker and Bassett (1978)]. Theoretical advances on the behavior of the associated estimators have completed this process [Powell (1994), Chen, Linton and Van Keilegom (2003)]. In empirics, partly thans to the generalization to quantile regressions, new fields of potential applications were born [see Buchinsky (1994) for an example and Koenker and Hallock (2000), Buchinsky (2003), for a review]. The recent and fast development of computer technology has clearly to do with a new regard on these robust, but formerly viewed as too cumbersome, methods.

Linear median regression assumes a linear relation between the dependent variable y and the explanatory variables x. Only a null median assumption is imposed on the disturbance process. Such a condition of identification "by the median" is entirely in agreement with the nonparametric literature. Since Bahadur and Savage (1956), that is known that without strong distributional assumptions such as normality, it is impossible to perform powerful tests on the mean of $i.i.d.$ observations, for any sample size. Moments are not empirically meaningful. That form of non-identification can be eliminated, even in finite samples, by choosing another measure of central tendency, such as the median. Hypotheses on the median value of non-normal observations can easily be tested by signs tests [see Pratt and Gibbons (1981)]. For setups that involve nonnormality, one may expect models with median identification to be more promising than their mean counterpart.

Median regression (and related quantile regressions) constitutes a good intermediate between parametric and nonparametric models. Distributional assumptions on the disturbance process are relaxed but the functional form stays parametric. Associated estimators are more robust to outliers than usual LS methods and may be more efficient whenever the median is a better measure of location than the mean. This holds for heavy-tailed distributions or distributions that have mass at 0. They are especially appropriate when unobserved heterogeneity is suspected in the data.[1] The most studied estimator is the Least Absolute Deviations (LAD) estimator. The current expansion of such "semiparametric" techniques [see Powell (1994)] reflects an intention to depart from restrictive parametric framework. However, related inference and confidence intervals remain based on asymptotic normality approximations. One may find disappointing this reversal to normal approximate inference when so much effort has been made to get rid of the strictly parametric model.

In this paper, we show that a testing theory based on signs of residuals provides a entire system of finite-sample exact inference for a linear median regression model. Exact tests and confidence regions remain valid under general assumptions involving heteroskedasticity of unknown form and nonlinear dependence. The level of these tests corresponds to the nominal level, for any sample

[1] The reader is referred to Buchinsky (1994) for an interpretation in terms of inequality and mobility topics in the US labor market, Engle and Manganelli (2000) for an application in Value at Risk issues in finance and Koenker and Hallock (2000) for an exhaustive review of this literature.

size.

The starting point is a well known result of quasi-impossibility in the statistical literature. Lehmann and Stein (1949) proved that sign methods were the only possible way of producing inference procedures under conditions of heteroskedasticity of unknown form when the number of observations is finite. All other inference methods, including HAC methods, that are not based on signs, are invalid for any sample size. These results had been used to derive and validate nonparametric sign tests but were barely exploited in econometrics. Our point is to stress their robustness and to generalize their use for median regression.

To our knowledge, sign-based methods have not received much interest in econometrics, compared to ranks or signed ranks methods. Dufour (1981), Campbell and Dufour (1991, 1995), Wright (2000), derive exact nonparametric tests for different time series models. In a regression context, Boldin, Simonova and Tyurin (1997) develop inference and estimation for linear models. They present both exact and asymptotic-based inferences for $i.i.d$ observations, whereas for autoregressive processes with $i.i.d$ disturbances, only asymptotic justification is available. Our work is positioned in the following of Boldin et al. (1997). We keep sign-statistics related to locally optimal sign tests that present a simple quadratic form and can easily be adapted for estimation. However, we extend their distribution-free properties to some dependent scheme in the data. We propose to conjugate them with projection techniques and Monte Carlo tests to systematically derive exact confidence sets.

The pivotality of the sign-based statistics validates the use of Monte-Carlo tests [Dwass (1957), Barnard (1963) and Dufour (2002)]. The Monte-Carlo method, adapted to discrete statistics by a tie-breaking procedure [Dufour (2002)], permits to obtain exact simultaneous confidence region for β. Then, conservative confidence intervals for each component of the parameter (or any real funtion of the parameter) are obtained by projection [Dufour and Kiviet (1998), Dufour and Taamouti (2000), Dufour and Jasiak (2001)]. Exact CI may not be bounded for certain non identifiable component. That results from the exactness of the method that insures the true value of the component belongs to exact CI with probability higher than $1 - \alpha$. In practice, computation of estimators and bounds of each confidence intervals requires global optimization algorithms [Goffe, Ferrier and Rogers (1994)].

Sign-based inference methods constitute an alternative to inference derived from the asymptotic behavior of the well known Least Absolute Deviations (LAD) estimator. The LAD estimator (such as related quantile estimators) is proved to be consistent and asymptotically normal in case of heteroskedasticity [Powell (1984) and Zhao (2001) for efficient weighted LAD estimator], or temporal dependence [Weiss (1991)]. Fitzenberger (1997*b*) extends the scheme of potential temporal dependence including ARMA disturbance processes. Horowitz (1998) uses a smoothed version of LAD. At the same time, an important advance in the LAD literature has been to provide good estimates of the asymptotic covariance matrix, which inference relies on. Powell (1984) proposes kernel estimation, but the most widespread method of estimation is the bootstrap [Buchinsky (1995)]. Buchinsky (1995) advocates the use of design matrix bootstrap for independent observations. In dependent cases, Fitzenberger (1997*b*) proposes moving block bootstrap. Finally, Hahn (1997) presents Bayesian bootstrap. The reader is referred to Buchinsky (1995, 2003), for a review and to Fitzenberger (1997*b*) for a comparison between these methods. Basically, kernel estimation is sensitive to the choice of the kernel function and of the bandwidth parameter, and the estimation of the LAD asymptotic covariance matrix needs a reliable estimator of the error term density at zero, which may be tricky especially for heteroskedastic disturbances. Besides, whenever the normal distribution is not a good finite-sample approximation, inference based on covariance matrix estimation may be problematic. From a finite-sample point of view, asymptotically justified methods can be arbitrarily unreliable. Coverage levels can be far from nominal ones. One can find examples of such distortions for time series context in Dufour (1981), Campbell and Dufour (1995, 1997) and for L_1-estimation in Buchinsky (1995), De Angelis, Hall and Young (1993), Dielman and Pfaffenberger (1988*a*, 1988*b*). Thus, inference based on signs constitutes a potential alternative when asymptotics fails. It remains to show that this alternative is performing. Other notable areas of investigation in the L_1 literature, gather study of nonlinear functional forms and structural models with endogeneity ["censored quantile regressions", Powell (1984, 1986) and Fitzenberger (1997*a*), Buchinsky (1998), "simultaneous equations", Amemiya (1982), Hong and Tamer (2003)]. More recently, authors have been interested in allowing for misspecification [Kim and White (2002), Komunjer (2003), Jung (1996)].

We study now a linear median regression model where the (possibly dependent) disturbance process is assumed to have a null median conditional on some exogenous explanatory variables and its own past. This setup covers both standard conditional heteroskedasticity and dependent effect in the residuals variance (like in ARCH, GARCH, stochastic volatility models,...) but do not allow lagged dependent nor autocorrelation in the residuals. We first treat the problem of inference and show that pivotal statistics based on the signs of the residuals are available for any sample size. Hence, exact inference and exact simultaneous confidence region on β can be derived using Monte-Carlo tests. For more general processes that may involve stationary ARMA disturbances, these statistics are no longer pivotal. The asymptotic covariance matrix constitutes a nuisance parameter that may invalid inference. However, transforming sign-based statistics with standard HAC methods [see White (1980), Newey and West (1987), Andrews (1991)] allows to asymptotically get rid of nuisance parameters. We thus extend the validity of the Monte Carlo method. For these kinds of processes, we loose the exactness but keep an asymptotic validity. In particular, this asymptotic validity requires less assumptions on moments or density existence than usual asymptotic-based inference. Besides, it does not demand to evaluate the disturbance density at zero, which constitutes one of the major limitations of kernel methods. In practice, we derive sign-based statistics from locally most powerful test statistics. We obtain exact confidence intervals for each component or any real function of β by projection techniques [see Dufour (2002)]. In a companion paper, we derive sign-based estimators that optimize these statistics and study their properties and performance [see Coudin and Dufour (2005)]. Once again, we insist on the fact that the use of sign-based statistics provides finite-sample valid inference which is not the case for usual inference theories associated with LAD and other quantile estimators that are based on their asymptotic behaviors.

The remainder of the paper is composed as follows. In section 2, we present model and notations. Section 3 gathers general results on exact inference. They are applied to median regressions in section 4. In section 5, we derive conservative confidence intervals at any given confidence level and illustrate the method on a numerical example. Section 6 is dedicated to the asymptotic validity of the finite-sample based inference method. In section 7, we give simulation results from comparisons to usual techniques. Section 8 concludes.

2. Framework

2.1. Model

We consider a stochastic process $W = \{W_t = (y_t,\ x_t') : \Omega \to \mathbb{R}^{p+1},\ t = 1, \ldots, n\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$. Let $\{\mathcal{F}_t,\ t = 1, \ldots, n\}$ be an adapted nondecreasing sequence of sub σ-fields of $\mathcal{F}$, i.e. $\mathcal{F}_t$ is a σ-field in Ω such that $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s < t$ and $\sigma(W_1, \ldots, W_t) \subset \mathcal{F}_t$, where $\sigma(W_1, \ldots, W_t)$ is the σ-algebra spanned by $W_1, \ldots, W_t$. $W_t = (y_t, x_t')$, where y_t is the dependent variable and $x_t = (x_{t1}, \ldots, x_{tp})'$, a p-vector of explanatory variables .
We assume that y_t and x_t satisfy a linear model and we shall impose in the following some conditions on the median of the disturbance process:

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

or, in vector notation,

$$y = X\beta + u, \tag{2.2}$$

where $y \in \mathbb{R}^n$ is a vector of dependent variables, $X = [x_1, \ldots, x_n]'$ is an $n \times p$ matrix of explanatory variables, $\beta \in \mathbb{R}^p$ is a vector of parameters, and $u \in \mathbb{R}^n$ is a disturbance vector such that

$$u_t|(x_1, \ldots, x_n) \sim F_t(.|x_1, \ldots, x_n),\ \forall t.$$

In the classical linear regression framework, the u process is assumed to be a martingale difference with respect to $\mathcal{F}_t = \sigma(W_1, \ldots, W_t),\ t = 1, 2, \ldots$.

Definition 2.1 MARTINGALE DIFFERENCE. *Let $\{u_t, \mathcal{F}_t :\ t = 1, 2, \ldots\}$ be an adapted stochastic sequence. Then $\{u_t, t = 1, 2, \ldots\}$ is a martingale difference sequence with respect to $\{\mathcal{F}_t, t = 1, 2, \ldots\}$ iff*

$$E(u_t|\mathcal{F}_{t-1}) = 0, \quad \forall t \geq 1.$$

We depart from this usual assumption. Indeed, our aim is to develop a framework that is robust to heteroskedasticity of unknown form. From Bahadur and Savage (1956), that is known that inference on the mean of $i.i.d$ observations of a random variable without any further assumption on the form of its distribution is impossible. Such a test has no power. This problem of non-testability can be viewed as a form of non-identification in a wide sense. Unless relatively strong distributional assumptions are made, moments are not empirically meaningful. Thus, if one wants to relax the distributional assumptions, one must choose another measure of central tendency such as the median. Such a measure is in particular well adapted if the distribution of the disturbance process does not possess moments.

As a consequence, in this median regression framework, martingale difference assumption can be replaced by an analogue in terms of median. We define the median-martingale difference or shortly said, mediangale that can be stated unconditional or conditional on the design matrix X.

Definition 2.2 STRICT MEDIANGALE. *Let* $\{u_t, \mathcal{F}_t,\ t = 1, 2\ldots\}$ *be an adapted sequence. Then* $\{u_t,\ t = 1, 2, \ldots\}$ *is a strict mediangale with respect to* $\{\mathcal{F}_t,\ t = 1, 2, \ldots\}$ *iff*

$$P[u_1 < 0] = P[u_1 > 0] = 0.5,$$

$$P[u_t < 0|\mathcal{F}_{t-1}] = P[u_t > 0|\mathcal{F}_{t-1}] = 0.5, \text{ for } t > 1.$$

Definition 2.3 STRICT CONDITIONAL MEDIANGALE. *Let* $\{u_t, \mathcal{F}_t,\ t = 1, 2\ldots\}$ *be an adapted sequence and* $\mathcal{F}_t = \sigma(u_1, \ldots, u_t, X)$*. Then* $\{u_t,\ t = 1, 2, \ldots\}$ *is a strict mediangale conditional on* X *with respect to* $\{\mathcal{F}_t,\ t = 1, 2, \ldots\}$ *iff*

$$P[u_1 < 0|X] = P[u_1 > 0|X] = 0.5,$$

$$P[u_t < 0|u_1, \ldots, u_{t-1}, X] = P[u_t > 0|u_1, \ldots, u_{t-1}, X] = 0.5, \text{ for } t > 1.$$

Let us define the **sign operator**: for $x \in \mathbb{R}$,

$$s(x) = \begin{cases} 1\,, & \text{if } x > 0 \\ 0\,, & \text{if } x = 0 \\ -1\,, & \text{if } x < 0. \end{cases}$$

Definition **2.3** may be generalized to cover the case where the errors may have a probability mass at zero.

Definition 2.4 WEAK CONDITIONAL MEDIANGALE. *Let* $\{u_t, \mathcal{F}_t,\ t = 1, 2\ldots\}$ *be an adapted sequence and* $\mathcal{F}_t = \sigma(u_1, \ldots, u_t, X)$*. Then* $\{u_t,\ t = 1, 2, \ldots\}$ *is a weak mediangale conditional on* X *with respect to* $\{\mathcal{F}_t,\ t = 1, 2, \ldots\}$ *iff*

$$P[u_1 > 0|X] = P[u_1 < 0|X],$$

$$P[u_t > 0|u_1, \ldots, u_{t-1}, X] = P[u_t < 0|u_1, \ldots, u_{t-1}, X],\ for\ t = 2, \ldots, n.$$

Stating that $\{u_t,\ t = 1, 2, \ldots\}$ is a weak mediangale with respect to $\{\mathcal{F}_t,\ t = 1, 2, \ldots\}$ is exactly equivalent to assuming that $\{s(u_t),\ t = 1, 2, \ldots\}$ is a difference of martingale with respect to the same sequence of sub-σ algebras $\{\mathcal{F}_t,\ t = 1, 2, \ldots\}$. However, the weak mediangale concept is slightly more demanding than usual statement of martingale difference of signs. Indeed, the sequence sub-σ algebras of reference is usually taken to $\{\mathcal{F}_t = \sigma(W_1, \ldots, W_t),\ t = 1, 2, \ldots\}$ and $\{s(u_t) \otimes x_t, \mathcal{F}_t\}$ is assumed to be martingale differences with respect to $\{\sigma(W_1, \ldots, W_t), ,\ t = 1, 2, \ldots\}$. Here, the sequence sub-σ algebras of reference is $\{\mathcal{F}_t = \sigma(W_1, \ldots, W_t, X),\ t = 1, 2, \ldots\}$. Conditional mediangale requires conditioning on the whole process X. We shall see later that asymptotic inference may be available under weaker assumptions, as a martingale difference on signs or more generally some mixing concepts on $\{s(u_t), \sigma(W_1, \ldots, W_t),\ t = 1, 2, \ldots\}$.

The conditional mediangale concept will be useful for developing exact inference (conditional on X).
We have replaced the difference of martingale assumption on the raw process u by a quasi-similar hypothesis on a robust transform of this process $s(u)$. Below we shall see it is relatively easy to deal with a weak mediangale by a simple transformation of the sign operator. But in order to simplify the presentation, we shall focus on strict mediangale concept. Therefore, the identification of our model will rely on the following assumption.

Assumption 2.5 CONDITIONAL STRICT MEDIANGALE. *The components of* $u = (u_1, \ldots, u_n)$ *satisfy a strict mediangale conditional on* X.

It is easy to see that Assumption **2.5** entails

$$\mathrm{med}(u_1|x_1, \ldots, x_n) = 0,$$
$$\mathrm{med}(u_t|x_1, \ldots, x_n, u_1, \ldots, u_{t-1}) = 0,\ t = 2, \ldots, n,$$
$$P[u_1 = 0|x_1, \ldots, x_n] = P[u_t = 0|u_1, \ldots, u_{t-1}, x_1, \ldots, x_n] = 0,\ t = 2, \ldots, n.$$

Hence, we really use median regression set-up.
Our last remark concerns exogeneity. As long as the x_t's are strictly exogenous explanatory variables, conditional mediangale concept is equivalent to usual martingale difference for signs with respect to $\mathcal{F}_t = \sigma(W_1, \ldots, W_t),\ t = 1, 2, \ldots$.

Proposition 2.6 MEDIANGALE EXOGENEITY. *Suppose* $\{x_t :\ t = 1, \ldots, n\}$ *is a strictly exogenous process for* β *and*

$$P[u_1 > 0] = P[u_1 < 0] = 0.5,$$
$$P[u_t > 0|u_1, \ldots, u_{t-1}, x_1, \ldots, x_t] = P[u_t < 0|u_1, \ldots, u_{t-1}, x_1, \ldots, x_t] = 0.5.$$

Then $\{u_t,\ t \in \mathbb{N}\}$ *is a strict mediangale conditional on* X.

To prove Proposition **2.6**, we use the fact that, as X is exogenous, $\{u_t,\ t \in \mathbb{N}\}$ does not Granger cause X.

Model (2.1) with the Assumption **2.5** allows for very general forms of the disturbance distribution, including asymmetric, heteroskedastic or dependent ones, as long as conditional medians are 0. We stress that neither density nor moment existence are required, which is an important difference with asymptotic theory. Indeed, what the mediangale concept requires is a form of independence in the signs of the residuals. This type of condition is in a sense similar to the CD condition for LAD regressions in Weiss (1991). It also extends results in Campbell and Dufour (1995, 1997). In the case of LAD regressions, Weiss uses it as part of the identification conditions required for showing consistency. Here, we focus on valid finite-sample inference without any further assumption on the form of the distributions. Since Fitzenberger (1997*b*), that is indeed known that weaker assumptions than Weiss CD condition for consistency of LAD (or quantile) estimator exist. However, consistency and asymptotic normality require more assumptions on moments.[2] With such a choice, testing theory is necessarily based on approximations (asymptotic or bootstrap). In order to conduct a fully exact method, we have to consider Assumption **2.5**.

2.2. Special cases

The above framework obviously covers independence but also a large spectrum of heteroskedasticity and dependence patterns. For example, suppose that

$$u_t = \sigma_t(x_1, \ldots, x_n)\ \varepsilon_t,\ t = 1, \ldots, n,$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are $i.i.d$ conditional on $X = [x_1, \ldots, x_n]'$. More generally, many dependence schemes are also covered: for example, any model of the form

$$\begin{aligned} u_1 = \quad & \sigma_1(x_1, \ldots, x_{t-1})\varepsilon_1\,, \\ u_t = \quad & \sigma_t(x_1, \ldots, x_{t-1}\,, u_1,\, \ldots\,,\, u_{t-1})\varepsilon_t\,, \quad t = 2,\, \ldots\,,\, n \\ & \text{where } \varepsilon_1,\, \ldots\,,\, \varepsilon_n \text{ are independent with median 0} \\ \sigma_1(x_1, \ldots, x_{t-1}) \text{ and} \quad & \sigma_t(x_1, \ldots, x_n\,, u_1,\, \ldots\,,\, u_{t-1})\,, \quad t = 2,\, \ldots\,,\, n\,, \\ \text{are non-zero with probability one.} & \end{aligned}$$

In time series context, this includes:

1. ARCH(q) with non-Gaussian noise ε_t, where

$$\sigma_t(x_1, \ldots, x_{t-1}\,, u_1,\, \ldots\,,\, u_{t-1})^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \cdots + \alpha_q u_{t-q}^2.$$

2. GARCH$(p,\, q)$ with non-Gaussian noises ε_t, where

$$\sigma_t(x_1, \ldots, x_{t-1}\,, u_1,\, \ldots\,,\, u_{t-1})^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \cdots + \alpha_q u_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2.$$

3. Stochastic volatility models with non-Gaussian noises ε_t where,

$$\begin{aligned} u_t &= \exp(w_t/2) r_y \varepsilon_t\,, \\ w_t &= a_1 w_{t-1} + \cdots + a_1 w_{t-p} + r_w v_t\,, \\ & \quad v_1,\, \ldots\,,\, v_n \text{ are. } i.i.d. \text{ random variables.} \end{aligned}$$

[2]In Fitzenberger (1997*b*), LAD and quantile estimators are shown to be consistent and asymptotically normal if amongst other, $E[x_t s_\theta(u_t)] = 0,\ \forall t = 1, \ldots, n.$, densities exist and second order moments for $(u_t,\ x_t)$ are finite.

The first example is especially relevant for cross-sectional data where procedures and estimators are expected to be robust to heteroskedasticity. Other examples present robustness properties to endogenous disturbances variance (or volatility) specification. Note again that the disturbance process is not required to be second order stationary. For non stationary processes that satisfy the mediangale Assumption, sign-based inference will work whereas all inference procedures based on asymptotic behaviorsof estimators will fail. Hence, they are more likely to be adapted to financial and macro applications. However, endogenous heterogeneity specifications are recently developed in other fields than finance [see Meghir and Pistaferri (2004) for an example in panel individual data].

Finally, the previous property is more general and does not specify explicitly the functional form of the variance as in an ARCH specification.

3. Exact finite-sample sign-based inference

3.1. Motivation

In econometrics, tests are often based on t or χ^2 statistics, which are derived from asymptotically normal statistics with a consistent estimator of the asymptotic covariance matrix. Unfortunately, in finite samples, these first order approximations can be very misleading. Confidence levels can be quite far from nominal ones: both the probability that an asymptotic test rejects a correct null hypothesis and the probability that a component of β can be contained in an asymptotic confidence interval may differ considerably from assigned nominal levels. One can find examples of such distortions in the literature, see for example Dufour (1981), Campbell and Dufour (1995, 1997), and for L_1 literature, Buchinsky (1995), De Angelis et al. (1993), Dielman and Pfaffenberger (1988*a*, 1988*b*). This remark usually motivates the use of bootstrap procedures. In a sense, bootstrapping (once bias corrected) is a way to make approximation closer by introducing artificial observations. However, the bootstrap still relies on approximations and in general there is no guarantee that the level condition can be satisfied in finite samples. Another way to look at the non-validity of asymptotics in finite samples is to recall a theorem proved by Lehmann and Stein [see Pratt and Gibbons (1981) and Lehmann and Stein (1949)]. Consider the very general hypothesis:

$$H_0 : \quad X_1, \ldots, X_n \text{ are independent observations each one with a distribution symmetric about zero.} \tag{3.1}$$

Here, H_0 allows for arbitrary heteroskedasticity. Let

$$\mathcal{H}_0 = \{F \in \mathcal{F}_n : F \text{ satisfies } H_0\}. \tag{3.2}$$

For this setup, Lehmann and Stein (1949) established the following theorem.

Theorem 3.1 *If a test has level α for H_0, where $0 \leq \alpha < 1$, then it must satisfy the condition*

$$\mathsf{P}[\, \textit{Rejecting } H_0 \mid |X_1|, \ldots, |X_n| \,] \leq \alpha \textit{ under } H_0 \,. \tag{3.3}$$

In other words, the only way to perform a test of H_0 must be by a sign test. Theorem **3.1** also implies that all procedures that do not satisfy condition 3.3 have size one. Note that procedures typically designated as "robust to heteroskedasticity" or "HAC" [see White (1980), Newey and West (1987), Andrews (1991), etc..], do not satisfy condition , consequently they have size one. On the other hand, sign-based procedures do satisfy this condition. Besides, as we will show in the next section, distribution-free pivotal sign-based statistics are available even in finite samples. They have been used in the statistical literature to derive nonparametric sign tests. The combination of both remarks give the theoretical basis for developing an exact inference method and a "robust" estimator. The most common procedure for developing inference on a statistical model can be described as follows. First, one finds an (hopefully consistent) estimator; second, the asymptotic distribution of the latter is established, from which confidence sets and tests are derived. Here, we shall proceed in the reverse order. We study first the test problem, then build confidence sets, and finally estimators. Hence, results on the valid finite-sample test problem will be adapted to obtain valid confidence intervals and estimation theory.[3]

3.2. Distribution-free pivotal functions and nonparametric tests

When the disturbance process is a conditional mediangale, the joint distribution of the signs of the disturbances is completely determined. These signs are mutually independent according to a uniform Bernoulli distribution on $\{-1,\, 1\}$. We state more precisely this result in the following proposition. We see also that the case with a mass at zero can also be covered provided a transformation in the sign operator definition.

Proposition 3.2 SIGN DISTRIBUTION. *Under model (*2.1*), suppose the errors* $(u_1, \ldots, u_n)$ *satisfy a strict mediangale conditional on* $X = [x_1, \ldots, x_n]'$. *Then the variables* $s(u_1), \ldots,\, s(u_n)$ *are i.i.d. conditional on* X *according to the distribution*

$$\mathsf{P}[s(u_t) = 1\,|x_1, \ldots, x_n] = \mathsf{P}[s(u_t) = -1\,|x_1, \ldots, x_n] = \frac{1}{2}\,, \quad t = 1,\, \ldots\,,\, n\,. \tag{3.4}$$

More generally, this result holds for any combination of $t = 1, \ldots, n$. If there is a permutation $\pi:\; i \rightarrow j$ such that mediangale property holds for j, the signs are $i.i.d.$

From the above proposition, it follows that the vector of the aligned signs

$$s(y - X\beta) = [s(y_1 - x_1'\beta),\, \ldots\,,\, s(y_n - x_n'\beta)]' \tag{3.5}$$

has a nuisance-parameter-free distribution (conditional on X), i.e. it is a **pivotal function**. Furthermore, any function of the form

$$T = T\big[s(y - X\beta),\, X\big] \tag{3.6}$$

is pivotal conditional on X. Indeed, $s(y - X\beta)$ is a vector of n independent Bernoulli components that take the value 1 with probability 0.5 and -1 with probability 0.5. Once the form of T is specified, the distribution of the statistic T is totally determined and can be simulated.

[3]For the estimation theory, the reader is referred to the companion paper Coudin and Dufour (2005).

Thanks to Proposition **3.2**, it is possible to construct tests for which the size is fully and exactly controlled. Consider testing

$$H_0(\beta_0): \ \beta = \beta_0 \text{ against } H_1(\beta_0): \beta \neq \beta_0.$$

Under H_0, $s(y_t - x_t'\beta_0) = s(u_t), \ t = 1, \ldots, n$. Thus, conditional on X,

$$T[s(y - \beta_0 X), X] \sim T(S_n, X) \tag{3.7}$$

where $S_n = (s_1, \ldots, s_n)$ and $s_1, \ldots, s_n$ are $i.i.d$ random variables according to a uniform Bernoulli distribution on $\{-1, \ 1\}$. If the distribution of $T(S_n, X)$ cannot be derived analytically, it can be easily simulated. A test with level α rejects the null hypothesis when

$$T[s(y - \beta_0 X), X] > c_T(X, \alpha) \tag{3.8}$$

where $c_T(X, \alpha)$ is the $(1-\alpha)$-quantile of the distribution of $T(S_n, X)$.

This method can be extended to error distributions with a mass at zero. Indeed, let us return to the case when:

$$\begin{gathered} \mathsf{P}[u_1 > 0 \,|\, X] = \mathsf{P}[u_1 < 0 \,|\, X], \\ \mathsf{P}[u_t > 0 \,|\, X, \ u_1, \ \ldots\ , \ u_{t-1}] = \mathsf{P}[u_t < 0 \,|\, X, \ u_1, \ \ldots\ , \ u_{t-1}], \quad t \geq 2\,. \end{gathered} \tag{3.9}$$

Besides dependence, this specification allows for discrete distributions with a probability mass at zero, i.e. we can have:

$$\mathsf{P}[u_t = 0 \,|\, X, \ u_1, \ \ldots\ , \ u_{t-1}] = p_t(X, \ u_1, \ \ldots\ , \ u_{t-1}) > 0 \tag{3.10}$$

where the $p_t(\cdot)$ are unknown and may vary between observations. An easy way out consists in modifying the sign function $s(x)$ as follows:

$$\tilde{s}(x, \, V) = s(x) + \big[1 - s(x)^2\big] s(V - 0.5), \quad \text{where} \quad V \sim U(0, \, 1), \tag{3.11}$$

or, equivalently,

$$\begin{aligned} \tilde{s}(x, \, V) \ &= s(x), && \text{if } s(x) = +1 \text{ or } -1 \\ &= s(V - 0.5), && \text{if } s(x) = 0. \end{aligned} \tag{3.12}$$

If V_t is independent of u_t then, irrespective of the distribution of u_t,

$$\mathsf{P}[\tilde{s}(u_t, \, V_t) = +1] = \mathsf{P}[\tilde{s}(u_t, \, V_t) = -1] = \frac{1}{2}\,. \tag{3.13}$$

Proposition 3.3 RANDOMIZED SIGN DISTRIBUTION. *Consider model (*2.1*) with the assumption that* $(u_1, \ \ldots\ , \ u_n)$ *constitute a weak mediangale conditional on* X*. Let* $V_1, \ \ldots\ , \ V_n$ *be i.i.d random variables following a* $U(0, \, 1)$ *distribution independent of* u *and* X*. Then the variables*

$\tilde{s}_t = \tilde{s}(u_t, V_t)$ are i.i.d. conditional on X according to the distribution

$$\mathsf{P}[\tilde{s}_t = 1 \mid X] = \mathsf{P}[\tilde{s}_t = -1 \mid X] = \frac{1}{2}, \quad t = 1, \ldots, n. \tag{3.14}$$

All the procedures described above can be applied without any further modification.

4. Regression sign-based tests

4.1. Regression sign statistics

The class of pivotal functions studied in the previous section is quite general. We must choose the form of the test statistic (the form of the T function) that provides the best power. To cover estimation, we will also need statistics that yield consistent tests for all values of β under "usual" identification hypotheses. Unfortunately, there is no uniformly most powerful test of $\beta = \beta_0$ against $\beta \neq \beta_0$. Hence, different alternatives may be considered. For testing $H_0(\beta_0) : \beta = \beta_0$ against $H_1(\beta_0) : \beta \neq \beta_0$ in model (2.1), we consider test statistics of the following form:

$$D_S(\beta_0, \Omega_n) = s(y - X\beta_0)' X \Omega_n[s(y - X\beta_0), X] X' s(y - X\beta_0) \tag{4.15}$$

where $\Omega_n[s(y - X\beta_0), X]$ is a $p \times p$ weight matrix that depends on the aligned signs $s(y - X\beta_0)$ under $H_0(\beta_0)$. Moreover, $\Omega_n[s(y - X\beta_0), X]$ is assumed to be positive definite.

Statistics associated with $\Omega_n = I_p$ and $\Omega_n = (X'X)^{-1}$ are given by

$$SB(\beta_0) = s(y - X\beta_0)' X X' s(y - X\beta_0) = \left\| X' s(y - X\beta_0) \right\|^2 \tag{4.16}$$

and

$$SF(\beta_0) = s(y - X\beta_0)' P(X) s(y - X\beta_0) = \left\| X' s(y - X\beta_0) \right\|_M^2 \tag{4.17}$$

where $P(X) = X(X'X)^{-1}X'$. Boldin et al. (1997) derive these statistics and show that they are associated with locally most powerful tests in case of $i.i.d$ disturbances under some regularity conditions on the distribution function. The locally most powerful test is well defined and unique when $dim(\beta) = 1$. When $\beta \in \mathbf{R}^p$ with $p > 1$, they choose the optimal test in term of maximization of the mean curvature, and obtain $SB(\beta_0)$. Their proof can easily be extended to disturbances that satisfy the mediangale property and for which the conditional density at zero is the same $f_t(0|X) = f(0|X),\ \forall t = 1, \ldots, n$. $SF(\beta_0)$ can be interpreted as a sign analogue of the Fisher statistic. More precisely, $SF(\beta_0)$ is a monotonic transformation of the Fisher statistic for testing $\gamma = 0$ in the regression of $s(y - X\beta_0)$ on X:

$$s(y - X\beta_0) = X\gamma + v. \tag{4.18}$$

Hence, $SF(\beta_0)$ may be interpreted as the Mahalanobis norm of $s(y - X\beta_0)$. These statistics are also related to Lagrange Multiplier (LM) type statistics. Koenker and Bassett (1982) develop Wald, LM and likelihood ratio asymptotic tests associated with the LAD estimator in a median regression.

They notice the advantage of the LM statistic that does not require to estimate the density of the disturbance at 0. In their framework, $D_S(\beta_0, \Omega_n)$ can be viewed as a LM test statistics of the simultaneous null hypothesis $H_0(\beta_0) : \beta = \beta_0$ for $i.i.d$ observations.
Note that if the disturbances are heteroskedastic, the locally optimal test statistic associated with the mean curvature will be of the following form.

Proposition 4.1 *In model (2.1), suppose the mediangale Assumption* **2.5** *holds, and the disturbances are heteroskedastic with conditional densities $f_i(.|X),\ i = 1, 2, \ldots$, that are C^1 around zero. Then, the locally optimal sign test statistic associated with the mean curvature is*

$$\tilde{SB}(\beta_0) = s(y - X\beta_0)'\tilde{X}\tilde{X}'s(y - X\beta_0) \tag{4.19}$$

where

$$\tilde{X} = \begin{Bmatrix} f_1(0|X) & 0 & \ldots \\ & f_i(0|X) & \\ 0 & \ldots & f_n(0|X) \end{Bmatrix} X.$$

However, when the $f_i(0|x)$'s are unknown, the optimal statistics is not reachable. The optimal weights must be replaced by approximations, such as weights derived from the normal distribution for instance.

In an estimation point of view, one may note that these test statistics can be interpreted as GMM statistics that exploit the property that $\{s_t \otimes x_t', \mathcal{F}_t\}$ is a martingale difference sequence. We saw in the first section that this property was induced by the mediangale Assumption **2.5**. However, these are quite unusual GMM statistics. Indeed, the parameter of interest is not defined by moment conditions in a explicit form. It is implicitly defined as the solution of some robust estimating equations (that involve aligned signs):

$$\sum_{t=1}^{n} s(y_t - x_t'\beta) \otimes x_t = 0$$

For $i.i.d$ disturbances, Godambe (2001) show that these estimating functions are optimal among all the linear unbiased (for the mean) estimating functions $\sum_{t=1}^{n} a_t(\beta)s(y_t - x_t'\beta)$. For independent and heteroskedastic disturbances, the set of optimal estimating equations is

$$\sum_{t=1}^{n} s(y_t - x_t'\beta) \otimes \tilde{x}_t = 0.$$

In those cases X and resp. $\tilde{X}$ can be seen as optimal instruments for the linear model.

For dependent processes, we propose to use a weighting matrix directly derived from the asymptotic covariance matrix of $\frac{1}{\sqrt{n}}s(y - X\beta_0) \otimes X$. Let us denote this asymptotic covariance matrix by

$J_n[s(y - X\beta_0), X]$. We consider

$$\Omega_n[s(y - X\beta_0), X] = \frac{1}{n}\hat{J}_n[s(y - X\beta_0), X]^{-1} \tag{4.20}$$

where $\hat{J}_n[s(y - X\beta_0), X]$ stands for a consistent estimate of $J_n[s(y - X\beta_0), X]$. This leads to

$$D_S(\beta_0, \frac{1}{n}\hat{J}_n^{-1}) = \frac{1}{n}s(y - X\beta_0)'X\hat{J}_n^{-1}X's(y - X\beta_0). \tag{4.21}$$

$J_n[s(y - X\beta_0), X]$ accounts for dependence among signs and explanatory variables. Hence, by using an estimate of its inverse as weighting matrix, we perform a HAC correction. More precisely, if the process $\{V_t(\beta_0) = s(y_t - x_t'\beta_0) \otimes x_t,\ t = 1, \ldots, n\}$ is second order stationary, we have

$$J_n = \sum_{j=-n+1}^{n-1} \Gamma_n(j), \tag{4.22}$$

where

$$\Gamma_n(j) = \begin{cases} \frac{1}{n}\sum_{t=j+1}^{n} E[V_t(\beta_0)V'_{t-j}(\beta_0)], & \text{for } j \geq 0, \\ \frac{1}{n}\sum_{t=-j+1}^{n} E[V_{t+j}(\beta_0)V'_t(\beta_0)], & \text{for } j < 0, \end{cases} \tag{4.23}$$

and the V_t process has spectral density matrix

$$f(\lambda) = \frac{1}{2\pi}\sum_{j=-\infty}^{+\infty} \Gamma(j)e^{-ij\lambda}, \text{ with } \Gamma(j) = E[V_tV'_{t-j}]. \tag{4.24}$$

J_n can be consistently estimated by kernel estimators of the spectral density matrix. The class of estimators we will consider was introduced by Parzen (1957) and corresponds to

$$\hat{J}_n = \frac{n}{n-p}\sum_{j=-n+1}^{n-1} k(\frac{j}{B_n})\hat{\Gamma}_n(j), \tag{4.25}$$

where

$$\hat{\Gamma}_n(j) = \begin{cases} \frac{1}{n}\sum_{t=j+1}^{n} V_t(\beta_0)V'_{t-j}(\beta_0) & \text{for } j \geq 0 \\ \frac{1}{n}\sum_{t=-j+1}^{n} V_{t+j}(\beta_0)V'_t(\beta_0) & \text{for } j < 0, \end{cases} \tag{4.26}$$

and $k(.)$ is a real-valued kernel. Modified kernel functions have been considered in the literature [see White (1984), Newey and West (1987), Andrews (1991)]. The bandwidth parameter B_n can be either fixed or automatically adjusted [see Andrews (1991)].

In all cases, the null hypothesis $\beta = \beta_0$ is rejected when the statistic evaluated at $\beta = \beta_0$ is large:

$$\begin{aligned} SB(\beta_0) &> c_{SB}(X, \alpha), \\ SF(\beta_0) &> c_{SF}(X, \alpha), \end{aligned}$$

$$D_S(\beta_0, \frac{1}{n}\hat{J}_n^{-1}) \quad > \quad c_{SHAC}(X,\, \alpha),$$

where $c_{SB}(X,\, \alpha)$, $c_{SF}(X,\, \alpha)$ and $c_{SHAC}(X,\, \alpha)$ are critical values which depend on the level α of the test. Since we are looking at pivotal functions, the critical values can be evaluated to any degree of precision by simulation. But a more elegant solution consists in using the technique of **Monte Carlo tests**, which can be viewed as a finite-sample version of the **bootstrap**.

4.2. Monte Carlo tests

Monte Carlo tests have been introduced by Dwass (1957) and Barnard (1963) and can be adapted to all pivotal statistics, when the distributions can be simulated. For a general review and extensions in the case of the presence of a nuisance parameter, the reader is referred to Dufour (2002). In our case, all previous tests are on the same model: given a statistic T, the test rejects the null hypothesis when T is large, i.e. when $T \geq c$, where c depends on the level of the test. Moreover, the conditional distribution of T given X is free of nuisance parameters. All ingredients are present to apply Monte Carlo testing procedure. We denote by $G(x) = P[T \geq x]$ the survival function and by $F(x) = P[T \leq x]$ the distribution function.
Let T_0 be the observed value of T and $T_1, \ldots, T_N$, N independent replicates of T. The empirical p-value is given by

$$\hat{p}_N(x) = \frac{N\hat{G}_N(x) + 1}{N+1} \tag{4.27}$$

where

$$\hat{G}_N(x) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{[0,\infty)}(T_i - x), \quad \mathbf{1}_A(x) = \begin{cases} 1, \text{ if } x \in A, \\ 0, \text{ if } x \notin A. \end{cases}$$

Then we have

$$P[\hat{p}_N(T_0) \leq \alpha] = \frac{I[\alpha(N+1)]}{N+1}, \text{ for } 0 \leq \alpha \leq 1,$$

where $I[x]$ stands for the largest integer less than equal to x; see Dufour (2002). If N is such that $\alpha(N+1)$ is an integer, then the randomized critical region has the same size as the usual critical region $\{G(T_0) \leq \alpha\}$.

In the case of **discrete distributions**, the method must be adapted to deal with ties. Indeed, the usual order relation on $\mathbb{R}$ is not appropriate for comparing discrete realizations that have a strictly positive probability to be equal. Different procedures have been presented in the literature to decide what to do when ties occur. They can be classified between random and non random procedures, both aiming to exactly control back the level of the test. For a good review of this problem, the reader is referred to Coakley and Heise (1996). For evaluating empirical survival functions in case of discrete statistics, Dufour (2002) advocates the use of a random tie-breaking procedure that relies on replacing the usual order relation by a lexicographic order relation, which is complete for discrete realizations. Therefore, each replication T_j is associated with a uniform independent W_j to form the pairs $(T_j,\, W_j)$. Pairs are compared using the lexicographical order:

$$(T_i,\ W_i) \geq (T_j,\ W_j) \Leftrightarrow \{T_i > T_j\ or\ (T_i = T_j\ and\ W_i \geq W_j)\}.$$

The adapted empirical p-value is given by

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

where

$$\tilde{G}_N(x) = 1 - \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{[0,\infty)}(x - T_i) + \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{[0]}(T_i - x)\mathbf{1}_{[0,\infty)}(W_i - W_0).$$

Then

$$P[\tilde{p}_N(T_0) \leq \alpha] = \frac{I[\alpha(N+1)]}{N+1},\ for\ 0 \leq \alpha \leq 1.$$

This random tie-breaking allows one to exactly control the level of the procedure. This also increases the power of the test.

Here, we consider testing in (2.1), $H_0(\beta_0)$: $\beta = \beta_0$ against $H_1(\beta_0)$: $\beta \neq \beta_0$ at level α, under a mediangale assumption on the errors using a statistics of the form $DS(\beta, \Omega_n)$. Take for example, $SF(\beta)$. After computing $SF(\beta_0)$ from the data, we choose N the number of replicates, such that $\alpha(N+1)$ is an integer, where α is the desired level. Then, we generate N replicates $SF_j = S_j' X(X'X)^{-1}X'S_j$ where S_j is a realization of a n-vector of independent Bernoulli random variables, and we compute $\tilde{p}_N(SF(\beta_0))$. Finally, $H_0(\beta_0)$ is rejected at level $1-\alpha$ if $\tilde{p}_N(SF(\beta_0)) < \alpha$.

5. Regression sign-based confidence sets and confidence distributions

In the previous section, we have shown how to obtain a sign-based test of $H_0(\beta_0)$: $\beta = \beta_0$ against $H_1(\beta_0) : \beta \neq \beta_0$, for which we can exactly control the level for any given finite number of observations. In this section, we discuss how to use such tests in order to build confidence sets for β with known level. The basic idea is the following one. For each value $\beta_0 \in \mathbb{R}^p$, do the Monte Carlo sign test $\beta = \beta_0$ against $\beta \neq \beta_0$ and get the associated simulated p-value. This p-value can be viewed as the "degree of confidence" one may have in $\beta = \beta_0$. As we will see later, that idea is clearly related to the notion of **confidence distribution** when β is of dimension one [see Schweder and Hjort (2002)].

The confidence set with level $1-\alpha$, $C_\beta(\alpha)$ is the set of all β_0 with p-value higher than $1-\alpha$. By construction, $C_\beta(\alpha)$ has level $1-\alpha$. From this simultaneous confidence set for β, it is possible, by **projection techniques**, to derive confidence intervals for the components. More generally, we can obtain conservative confidence sets for any transformation $g(\beta)$ where g can be any kind of real function, including nonlinear ones. Obviously, obtaining a continuous grid of $\mathbb{R}^p$ is not realistic.

We will instead require **global optimization search algorithms**.

5.1. Confidence sets and conservative confidence intervals

Projection techniques yield finite-sample conservative confidence intervals and confidence sets for general functions of the parameter β. For examples of use in different settings and for further discussion, the reader is referred to Dufour (1990), Dufour (1997), Abdelkhalek and Dufour (1998), Dufour and Kiviet (1998), Dufour and Jasiak (2001), Dufour and Taamouti (2000). The basic idea is the following one. Suppose a simultaneous confidence set with level $1-\alpha$ for the entire parameter β is available. Its orthogonal projections on each axis of $\mathbb{R}^p$ will give confidence intervals for each component β_k of size larger than $1-\alpha$. To be more precise, suppose $C_\beta(\alpha)$ is a confidence set with level $1-\alpha$ for β :

$$\mathsf{P}[\beta \in C_\beta(\alpha)] \geq 1-\alpha. \tag{5.1}$$

Since

$$\beta \in C_\beta(\alpha) \Longrightarrow g(\beta) \in g\big(C_\beta(\alpha)\big), \tag{5.2}$$

we have:

$$\mathsf{P}[\beta \in C_\beta(\alpha)] \geq 1-\alpha \Longrightarrow \mathsf{P}[g(\beta) \in g\big(C_\beta(\alpha)\big)] \geq 1-\alpha\,.$$

Thus, $g(C_\beta(\alpha))$ is a conservative confidence set for $g(\beta)$. If $g(\beta)$ is scalar, the interval (in the extended real numbers)

$$I_g[C_\beta(\alpha)] = \left[\inf_{\beta \in C_\alpha(\beta)} g(\beta)\,,\ \sup_{\beta \in C_\alpha(\beta)} g(\beta)\right]$$

also has level $1-\alpha$:

$$\mathsf{P}\left[\min_{\beta \in C_\alpha(\beta)} g(\beta)\ \leq g(\beta) \leq \max_{\beta \in C_\alpha(\beta)} g(\beta)\right] \geqslant 1-\alpha. \tag{5.3}$$

Hence, to obtain valid conservative confidence intervals for the component β_k of the β parameter in the model (2.1) under mediangale Assumption **2.5**, it is sufficient to solve both following numerical optimization problems, (stated here for the statistic SF),

$$\begin{aligned} &\min_{\beta \in \mathbb{R}^p}\ \beta_k \\ &\text{s.t. }\ \tilde{p}_N(SF(\beta)) \geq 1-\alpha \end{aligned}$$

and

$$\begin{aligned} &\max_{\beta \in \mathbb{R}^p}\ \beta_k \\ &\text{s.t. }\ \tilde{p}_N(SF(\beta)) \geq 1-\alpha \end{aligned}$$

where $\tilde{p}_N$ is constructed as proposed in the previous section, thanks to N replicates SF_j of statistic SF under the null. This can be done easily in practice by using a global search optimization algorithm, like **simulated annealing algorithm** [Press, Teukolsky, Vetterling and Flannery (2002) and Goffe et al. (1994)].

5.2. Confidence distribution

As we saw in the last section, we can associate a simulated p-value to each possible value β_0 of β. This p-value can be seen as a sort of degree of confidence one may have in the value β_0. By aggregating values of β with given p-values, we can construct simultaneous confidence regions for any significance level of coverage.

In the special case when β has dimension one, the function that associates the simulated p-value to each value β_0 of β is related to a **confidence distribution** of β given the realizations $(y,\ X)$ [Schweder and Hjort (2002)]. The confidence distribution is essentially the same idea as the Fisher fiducial distribution. Its quantiles span all possible confidence intervals. More precisely, the confidence distribution of β is defined as a distribution with cumulative $CD(\beta)$ and quantile function $CD^{-1}(\beta)$, such that

$$P_\beta[\beta \leq C^{-1}(\alpha; y; X)] = P_\beta[C(\beta; y; X) \leq \alpha) = \alpha \tag{5.4}$$

for all $\alpha \in (0,1)$ and **for all probability distributions in the statistical model.** Hence, $(-\infty,\ CD^{-1}(\alpha)]$ constitutes a one-sided stochastic confidence interval with coverage probability α.[4] The realized confidence $CD(\beta_0; y; X)$ is the p-value of the one-sided hypothesis $H_0 :\ \beta \leq \beta_0$ versus $H_1 :\ \beta > \beta_0$ when the observed data are $y,\ X$. Equivalently the realized p-value when testing $H_0 :\ \beta = \beta_0$ versus $H_1 :\ \beta \neq \beta_0$ is $2\min\{CD(\beta_0), 1 - CD(\beta_0)\}$. In case of discrete statistics, such as sign-based statistics $DS(\beta, \Omega)$, confidence distributions need an adaptation. Typically, half correction takes the form

$$CD(\beta) = P_\beta[DS > ds_{obs}] + \frac{1}{2}P_\beta[DS = ds_{obs}] \tag{5.5}$$

where DS is the test statistic and ds_{obs}, the observed value.
As the cumulative distribution $CD(\beta)$ is an invertible function of β and is uniformly distributed, $CD(\beta)$ constitutes a pivot conditional on X. Reciprocally, whenever a pivot is available for example any statistics $T(\beta)$ that increases with β and that has a cumulative distribution function F independent of β and of any nuisance parameter, $F[T(\beta)]$ is uniformly distributed and thus turns to be a confidence distribution. Hence, simulated p-values associated with sign statistics $DS(\beta, \Omega)$ shall lead to approximated confidence distribution.

Confidence distributions can be very useful tools. They summarize the results of inference on β. They can, in a sense, be compared to Bayesian posterior probabilities for a frequentist setup. To any value of β, they associate a valid "degree of confidence". This "degree of confidence" is also clearly related to the degree of identification of the parameter. For a non identified parameter, that will

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

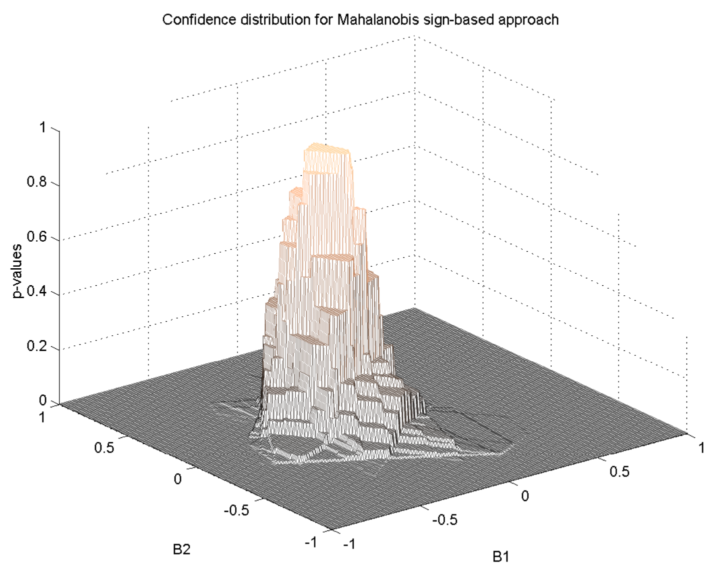
be impossible to obtain tight confidence intervals nor large p-values. Thus, we expect the p-value associated with the estimate to give us an idea on the degree of identification of the underlying parameter. Confidence distributions are not defined for $p \geq 2$. However, this is still possible to build p-values for some β of interest. In particular, the p-value associated with the estimates. See subsection 5.3 for an example.

5.3. Numerical illustration

This part reports a simulated example as a numerical illustration. We generate the following normal mixture process, for $n = 50$,

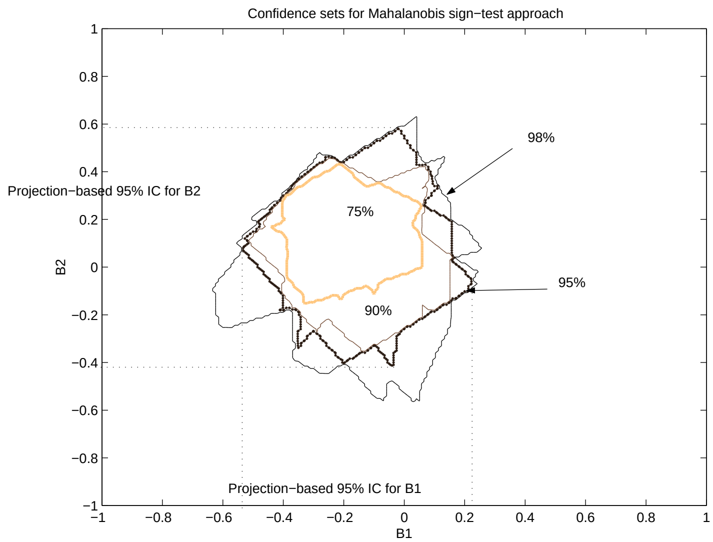
$$
\begin{aligned}
y_t &= \beta_0 + \beta_1 x_t + u_t\,, \quad t = 1, \ldots, n, \\
& u_t \overset{i.i.d.}{\sim} \begin{cases} N[0,\, 1] & \text{with probability } 0.95 \\ N[0,\, 100^2] & \text{with probability } 0.05 \end{cases} \\
& \beta_0 = \beta_1 = 0.
\end{aligned}
\tag{5.6}
$$

We conduct an exact inference procedure with 1000 replicates. We perform tests of $H_0 : \ \beta = \beta^*$ on a grid of values of $\beta^* = (\beta_0^*, \beta_1^*)$ and retain the associated simulated p-values. As β is a 2-vector, we can have a graphic illustration. To each value of the vector β is associated the corresponding simulated p-value. This leads to a 3 dimension graphics of confidence distribution. Confidence region with coverage level $1 - \alpha$ is obtained by gathering all values of β where p-values are bigger than $1 - \alpha$. Same graphics for other estimators are provided in appendix.



"Confidence distribution" of SF-based inference

The simulated p-value gives the "degree of confidence" one may have in the associated value of β. Valid confidence regions are obtained by gathering all values of β with p-value higher than the specified level. They turn to be the sections of the "confidence distribution".



Confidence regions provided by SF-based inference

Confidence region increases with the level of coverage and includes all other confidence regions with smaller level of coverage. They are highly non elliptic and thus very may lead to very different results than an asymptotic inference. To derive a conservative confidence interval, we just project the confidence region with the same specified level of coverage on the axis of β_0 or β_1.

In Table 5.3, we compare sign-based method to least squares estimators and asymptotic confidence intervals. As expected, LS estimation and inference are very sensitive to outliers whereas sign-based confidence sets appear to be largely more robust. The estimator associated with sign-based inference is the (or one of the) value(s) associated with the highest simulated p-value [Coudin and Dufour (2005)].

	OLS	White	SF
$\hat{\beta}_0$	-1.876	-1.876	-0.28
(s.d.)	*(1.343)*	*(1.293)*	-
90%CI	[-4.126, 0.375]	[-4.043, 0.292]	[-0.52, 0.17]
95%CI	[-4.574, 0.822]	[-4.474, 0.722]	[-0.54, 0.23]
98%CI	[-5.103, 1.351]	[-4.983, 1.232]	[-0.64, 0.26]
$\hat{\beta}_1$	0.358	0.358	0.14
(s.d.)	*(1.424)*	*(0.845)*	-
90%CI	[-2.030, 2.745]	[-1.058, 1.774]	[-0.36, 0.46]
95%CI	[-2.504, 3.219]	[-1.340, 2.055]	[-0.42, 0.59]
98%CI	[-3.065, 3.780]	[-1.673, 2.388]	[-0.57, 0.64]

For a more complete comparison of sign methods to usual ones including bootstrap and kernel estimation of asymptotic covariance matrix of LAD estimator, the reader is referred to section 7.

6. Asymptotic theory

This section is dedicated to asymptotic results. We point out that mediangale Assumption **2.5** can be seen as too restrictive and excludes some processes whereas usual asymptotic inference still can be conducted on these processes. We show that the finite-sample sign-based method of inference remains also asymptotically valid when Assumption **2.5** is relaxed. For a fixed number of replicates, when the number of observation goes to infinity, the level of a test tends to the nominal level. Besides, we stress the ability of our models to cover heavy-tailed distributions including infinite disturbance variance. This is roughly known in the literature concerning LAD estimation, but not enough exploited (in our opinion) in usual related conditions for consistency and asymptotic normality.

6.1. Asymptotic distributions of test statistics

In this part, we derive asymptotic distributions of sign statistics. We show that HAC correction of the test statistics $D_S(\beta_0, \frac{1}{n}\hat{J}_n^{-1})$ in (4.21) allows one to obtain asymptotically pivotal distribution. The set of assumptions we make to stabilize asymptotic behavior will be needed for further asymptotic results. We consider the test $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$ in the linear model (2.1), with the following assumptions.

Assumption 6.1 MIXING. *$\{(x_t', u_t)\}$ is α-mixing of size $-r/(r-2)$ with $r > 2$.*[5]

Assumption 6.2 MOMENT CONDITION. $E[x_t s(u_t)] = 0,\ \forall t = 1, \ldots, n,\ \forall n \in \mathbb{N}$.

Assumption 6.3 BOUNDEDNESS. *$x_t = (x_{1t}, \ldots, x_{pt})'$ and $E|x_{ht}|^r < \Delta < \infty,\ h = 1, \ldots, p,\ t = 1, \ldots, n,\ \forall n \in \mathbb{N}$.*

Assumption 6.4 NON-SINGULARITY. *$V_n = \mathrm{var}[\frac{1}{\sqrt{n}} \sum_{t=1}^{n} s(u_t)x_t]$ is uniformly positive definite.*

Assumption 6.5 CONSISTENT ESTIMATOR OF V_n. *$\Omega_n(\beta_0)$ is symmetric positive definite uniformly over n and $\Omega_n - \frac{1}{n}V_n^{-1} \to_p 0$.*

Theorem 6.6 ASYMPTOTIC DISTRIBUTION OF STATISTIC SHAC. *In model (2.1), with conditions* **6.1- 6.5***, we have, under H_0,*

$$D_S(\beta_0, \Omega_n) \to \chi^2(p).$$

[5]Mixing concept is defined in annex

Corollary 6.7 *In model (*2.1*), suppose the mediangale Assumption* **2.5** *and moment condition* **6.3** *are fulfilled. If $X'X/n$ is positive definite uniformly over n and converges in probability to a definite positive matrix, then, under H_0,*

$$SF(\beta_0) \to \chi^2(p).$$

When the mediangale condition holds, J_n reduces to $E(X'X/n)$. Hence, $(X'X/n)^{-1}$ is a consistent estimator of J_n^{-1}.

6.2. Asymptotic validity of Monte Carlo tests

We first state some general results on asymptotic validity of Monte-Carlo based inference methods. Then, we apply these results to sign-based inference method.

6.2.1. Generalities

Let us consider a parametric or semi parametric model $\{\mathrm{M}_\beta,\ \beta \in \Theta\}$, where the parameter β is identifiable. Let $S^n(\beta_0)$ be a test statistic of $H_0(\beta_0):\ \beta = \beta_0$ against $H_1(\beta_0):\ \beta \neq \beta_0$. Let $S_0^n(\beta_0)$ be the observed statistic and c_n the rate of convergence. We suppose that, under $H_0(\beta_0)$, $c_n S_0^n$ converges in law to a distribution $F(x)$ and we note $G(x)$ the corresponding survival function. We show in the following theorem that, if a series of conditional survival functions $\tilde{G}(x|X_n(\omega))$ given $X(\omega)$ satisfies

$$\tilde{G}_n(x|X_n(\omega)) \to G(x), \text{ with probability one,}$$

where G does not depend on the realization $X(\omega)$. The distribution of $c_n S_0^n(\beta_0)$ can be approximated by $\tilde{G}_n(x|X_n(\omega))$. Note that $G(x)$ can depend on some parameters of the distribution of X provided that it does not depend on realizations.

Theorem 6.8 GENERIC ASYMPTOTIC VALIDITY. *Let $S^n(\beta_0) = S_n^0$ be a test statistic for $H_0(\beta_0):\ \beta = \beta_0$ against $H_1(\beta_0):\ \beta \neq \beta_0$ in the model (*2.1*). Suppose that, under $H_0(\beta_0)$,*

$$P[c_n S_0^n \geq x|X_n] = G_n(x|X_n) = 1 - F_n(x|X_n) \to G(x)\ a.e.,$$

where $\{c_n\}$ is a sequence of positive constants and suppose that $\tilde{G}_n(x|X_n(\omega))$ is a series of survival functions such that

$$\tilde{G}_n(x|X_n(\omega)) \to G(x)\ a.e. \text{ with probability one.}$$

Then

$$\lim_{n\to\infty} P[\tilde{G}_n(c_n S_0^n, X_n(\omega)) \leq \alpha] \leq \alpha. \tag{6.1}$$

This theorem can also be stated in a Monte-Carlo based version. Following Dufour (2002), we define randomized empirical survival functions and randomized empirical p-values adapted to discrete statistics. Let $(U_0, U_1, \ldots, U_N)$ be a $N+1$ vector of $i.i.d$ real variables drawn from a $\mathcal{U}[0,1]$ distribution, S_0^n the observed statistics and $S^n(N) = (S_1^n, \ldots, S_N^n)$, N independent replicates drawn from $\tilde{F}_n(x)$. Then, the randomized empirical survival function under the null hypothesis is

$$\tilde{G}^r_{N,n}[x, n, U_0, S_0^n, S^n(N), U(N)] = 1 - \frac{1}{N}\sum_{j=1}^{N} u(x - c_n S_j^n) + \frac{1}{N}\sum_{j=1}^{N} \delta(c_n S_j^n - x) u(U_j - U_0) \tag{6.2}$$

with, $u(x) = 1_{[0,\,\infty)}(x)$, $\delta(x) = 1_{\{0\}}$. Note that $\tilde{G}^r_{N,n}[x, S_0^n, S^n(N), n]$ is in a sense an approximation of $\tilde{G}_n(x)$, and thus depends on N the number of replicates and n the number of observations. The randomized empirical p-value function is defined as

$$\tilde{p}^r_N(x) = \frac{N\tilde{G}^r_{N,n}(x) + 1}{N+1}. \tag{6.3}$$

We can now state the Monte Carlo-based version of Theorem **6.8**.

Theorem 6.9 MONTE CARLO TEST ASYMPTOTIC VALIDITY. *Let $S^n(\beta_0) = S_n^0$ be a test statistic for $H_0(\beta_0): \ \beta = \beta_0$ against $H_1(\beta_0): \ \beta \neq \beta_0$ in the model (*2.1*). Suppose that, under $H_0(\beta_0)$,*

$$P[c_n S_0^n \geq x | X_n] = G_n(x|X_n) = 1 - F_n(x|X_n) \to G(x) \ a.e.,$$

where $\{c_n\}$ is a sequence of positive constants. Let S_n be a random variable with conditional survival function $\tilde{G}_n(x|X_n)$ such that

$$P[c_n S_n \geq x | X_n] = \tilde{G}_n(x|X_n) = 1 - \tilde{F}_n(x|X_n) \to G(x) \ a.e.,$$

and $(S_1^n, \ldots, S_N^n)$ be a vector of N independent replication of S_n where $(N+1)\alpha$ is an integer. Then, the randomized version of the Monte Carlo test at level α is asymptotically valid, i.e.

$$\lim_{n\to\infty} P[\tilde{p}^r_N(\beta_0) \leq \alpha] \leq \alpha. \tag{6.4}$$

These results can be applied to sign-based inference method. However, Theorems **6.8** and **6.9** are much more general. They do not exclusively rely on asymptotic normality: the limiting distribution may be different from a Gaussian one. Besides, the rate of convergence may differ from $\sqrt{n}$.

6.2.2. Asymptotic validity of sign-based inference

In model (2.1), suppose that conditions **6.1- 6.5** hold and consider the testing problem

$$H_0(\beta_0) : \beta = \beta_0 \text{ against } H_1(\beta_0) : \beta \neq \beta_0.$$

Let $D_S(\beta, \hat{J}_n^{-1})$ be the test statistic as defined in (4.21).

- Observe $SF_0 = D_S(\beta_0, \hat{J}_n^{-1})$. Draw N replicates of sign vector as if the n observations were independent. The n components of the sign vectors are independent and drawn from a

$B(1, .5)$ distribution.

- Construct $(SF_1,\ \ SF_2, \ldots, SF_N)$, the N "pseudo" replicates of $D_S(\beta_0, X'X^{-1})$ under the null. We call them "pseudo" replicates because they are drawn as if observations were independent, which may be not the case for the observed $SF(\beta_0)$.

- Draw $N+1$ independent replicates $(W_0, \ldots, W_N)$ from a $\mathcal{U}_{[0,1]}$ distribution and form the couple $(SF^j,\ W_j)$, where SF^0 stands for the observed $SF(\beta_0)$.

- Compute $\tilde{p}_N^r(\beta_0)$ using (6.3).

- From Theorem **6.9**, the confidence region $\{\beta \in \mathbb{R}^p | \tilde{p}_N^r(\beta) \geq \alpha\}$ is asymptotically conservative with level at least $1-\alpha$. We reject $\mathcal{H}_0$ if $\tilde{p}_N^r(\beta_0) \leq \alpha$.

Remark that this method **does not require the existence of moments nor of a density on the** $\{u\}$ process. This differs from usual asymptotic tests. Indeed, usual asymptotically justified inference is based on the asymptotic behavior of estimators; see for example Fitzenberger (1997*b*) and Weiss (1991). Yet, this inference is very restrictive as the asymptotic behavior of the estimators relies largely on its asymptotic variance which may involve some unknown parameters such as the conditional density at 0 of the disturbance process u and that can be viewed as nuisance parameters. Their approximation and estimation constitute a large issue in inference field. This usually requires kernel methods. Here, we show that adopting our finite-sample sign-based inference, we get around some problems associated with asymptotic covariance matrix estimation.

7. Simulations and comparisons with other methods

We study the performance of sign-based methods compared to usual inference based on OLS or LAD estimators with different approximations for their asymptotic covariance matrices. We use the sign statistics $D_S[\beta, (X'X)^{-1}]$ and $D_S(\beta, \hat{J}_n^{-1})$ when a correction is needed for autocorrelation. We consider a set of general DGP to illustrate different classical problems one may encounter in practice. Results are presented in the way suggested by the theory. First, we investigate the performance of tests, then, confidence sets. We use the following linear regression model,

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

where $x_t = (1, x_{2,t}, x_{3,t})'$ and β_0 are 3×1 vectors. We denote the sample size T. Monte-Carlo studies are based on S generated random samples. We investigate the behavior of inference, confidence regions and estimators for 12 general DGP. For the first 7 ones, u_t's are $i.i.d$ or depend on the explanatory variables and their past values in a *multiplicative* heteroskedastic or dependent and stationary way,

$$u_t = h(x_t, u_{t-1}, \ldots, u_1)\epsilon_t. \tag{7.2}$$

In those cases, u constitutes a strict conditional mediangale given X (see Assumption **2.5**). Hence, the sign-based inference leads to exact levels. Next, we study the behavior of the sign-based inference (involving a HAC correction) when it is only asymptotically valid. In cases 8-9-10, x_t and u_t are such that $E(x_t u_t) = 0$ and $E[x_t sign(u_t)] = 0$. Finally, cases 11 and 12 illustrate two kinds of second order nonstationary disturbances. As we noted previously, sign-based inference does not require any stationary assumptions contrary to asymptotic tests derived from CLT. We thus expect sign-based tests to be the only procedures presenting some power.
For cases 1-2, 9-11, we employ the same stochastic processes as Fitzenberger (1997*b*).

$$
\begin{array}{ll}
\text{AR(1)-}\textit{HOM}\text{:} & x_{j,t} = \rho_x x_j, t-1 + \nu_{j,t},\ j = 2, 3. \\
 & u_t = \rho_u u_t + \nu_t^u, \\
\text{AR(1)-}\textit{HET}\text{:} & x_{j,t} = \rho_x x_{j,t-1} + \nu_{j,t}, \\
 & \tilde{u}_t = \rho_u \tilde{u}_t + \nu_t^u, \\
\text{and} & u_t = min\{3, max[0.21, |x_{2,t}|]\} \times \tilde{u}_t
\end{array}
$$

where the innovations ν_t^u and $\nu_{j,t}$ are centered normal with variance 1 and we let vary (ρ_x, ρ_u).

$$
\begin{array}{ll}
\text{CASE 1:} & \rho_x = \rho_u = 0,\ HOM. \\
\text{CASE 2:} & \rho_x = \rho_u = 0,\ HET. \\
\text{CASE 8:} & \rho_x = 0, \rho_u = 0.5,\ HOM. \\
\text{CASE 9:} & \rho_x = 0.5, \rho_u = 0.5,\ HET. \\
\text{CASE 10:} & \rho_x = 0, \rho_u = 0.9,\ HOM.
\end{array}
$$

For cases 3-8, we study other kinds of heteroskedasticity, nonlinear stationary dependence (GARCH, Stochastic Volatility ...), debalanced scheme in the explanatory variables, and heavy-tailed distributions for the error term.

CASE 3: Outlier

$$
y_t = x_t'\beta_0 + u_t,
$$

$$
u_t \overset{i.i.d.}{\sim} \begin{cases} N[0,\ 1] & \text{with probability } 0.95 \\ N[0,\ 1000^2] & \text{with probability } 0.05 \end{cases}
$$

CASE 4: Stationary GARCH(1,1)

$$
y_t = x_t'\beta_0 + u_t,
$$

$$
u_t = \sigma_t \epsilon_t,
$$

$$
\sigma_t^2 = 0.666 u_{t-1}^2 + 0.333 \sigma_{t-1}^2,
$$

where $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$.

CASE 5: Stochastic Volatility
$y_t = x_t'\beta_0 + u_t,$
$u_t = exp(w_t/2)\epsilon_t,$
$w_t = 0.5w_{t-1} + v_t,$
where $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, $v_t \overset{i.i.d.}{\sim} \chi_2(3)$.

CASE 6: Debalanced design matrix:
$y_t = x_t'\beta_0 + u_t,$
where $x_{2,t} \sim \mathcal{B}(1, 0.3),$
$x_{3,t} \overset{i.i.d.}{\sim} \mathcal{N}(0, .01^2),$
$u_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1).$

CASE 6 BIS: Deb. design matrix + het. disturbances
$y_t = x_t'\beta_0 + u_t,$
$x_{2t} \overset{i.i.d}{\sim} \mathcal{N}(0,1)$
$x_{3t} \overset{i.i.d}{\sim} \chi_2(1)$
$u_t = x_{3t}\epsilon_t,\ \epsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0,1)$

CASE 7: Cauchy disturbances
$y_t = x_t'\beta_0 + u_t,$
where $x_t \sim \mathcal{N}(0, I_2),$
$u_t \overset{i.i.d.}{\sim} \mathcal{C}.$

CASE 11: Non stationary GARCH(1,1)
$y_t = x_t'\beta_0 + u_t,$
$u_t = \sigma_t\epsilon_t,$
$\sigma_t^2 = 0.8u_{t-1}^2 + 0.8\sigma_{t-1}^2,$
where $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1).$

CASE 12: Exponential Variance
$y_t = x_t'\beta_0 + u_t,$
$u_t = exp(.2 * t)\epsilon_t,$
where $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1).$

To sum up, cases 1 and 2 present $i.i.d.$ normal observations without and with conditional heteroskedasticity. Cases 3 involves outliers in the error term. This can be seen as an example of measurement error in the observed y. Cases 4 and 5 are cases of nonlinear stationary dependent disturbances, when mediangale Assumption **2.5** holds. Case 6 presents a very debalanced scheme in the design matrix (a case when the LAD estimator is known to perform badly). Case 6 BIS combines non classical scheme in the design matrix and heteroskedasticity in the disturbances. Case 7 is an example of long tailed errors. Cases 8, 9 and 10 illustrate the behavior of sign-based inference when a linear dependence is present in the error term with different levels of autocorrelation. Finally, cases 11 and 12 involve disturbances that are not second-order stationary (nonstationary GARCH and exponential variance).

7.1. Size

We first study level distortion in small samples. We consider the testing problem

$$H_0[(1,2,3)'] : \beta_0 = (1,2,3)' \text{ against } H_1[(1,2,3)'] : \beta_0 \neq (1,2,3)'.$$

We compare tests based on $D_S[\beta, (X'X)^{-1}]$ and $D_S(\beta, \hat{J}_n^{-1})$ to various asymptotic Wald tests based on different estimates of the asymptotic covariance matrix of the LAD and of the OLS estimators. More precisely, we consider,

- IIDOLS: usual standard deviations of the OLS estimator under homoskedasticity and independence.
- WHOLS: usual White correction for heteroskedasticity of the OLS estimator.
- BARTOLS: Bartlett kernel estimator for OLS covariance matrix.
- OSLAD: order statistic estimator for asymptotic covariance matrix of LAD, (assume $i.i.d$ residuals, estimate of the residual density at 0 is obtained from a confidence interval constructed for the θth order statistics [see Buchinsky (2003)].
- BARTLAD: Bartlett kernel estimator for LAD asymptotic covariance matrix [see Powell (1984), Fitzenberger (1997*b*), Buchinsky (1995)] with automatic bandwidth parameter [see Andrews (1991)].
- DMBLAD: design matrix bootstrap centering around the sample estimate for the LAD estimator. This is a naive bootstrap version [see Buchinsky (2003)].
- MBBLAD: moving block bootstrap centering around sample estimate for the LAD estimator [see Fitzenberger (1997*b*)].
- SF: sign test based on the $D_S[\beta, (X'X)^{-1}]$ statistics.

- SHAC: sign test based on the $D_S(\beta, \hat{J}_n^{-1})$ statistics where $\hat{J}_n^{-1}$ is estimated by a Bartlett kernel.

In Table 8, we report the simulated level for a conditional test with nominal level $\alpha = 5\%$ given X. The number of replicates for the bootstrap and the sign methods is the same, i.e. 2999. All bootstrapped samples are of size $T = 50$. We perform $S = 5000$ simulations to estimate the level of these tests. For the two sign-statistics, we also report the asymptotic level of the test whenever the asymptotic distribution of the sign-statistic is easy to derive.

Cases when mediangale Condition 2.5 holds: when the mediangale condition holds, the finite-sample sign-based inference leads to exact levels.

Table 3:: Level of Conditional Tests of $H_0 : \beta = (1, 2, 3)'$.

$y_t = x_t\beta + u_t,\ t = 1, \ldots, T.$	*sign*		*LAD*				*OLS*		
$T = 50$	SF	HAC	OS	DMB	MBB5	BART	IID	WH	BART
***CASE 1**: $\rho_\epsilon = \rho_x = 0$, HOM.	.052 *.047***	.050 *.019***	.086	.050	.089	.047	.060	.096	.113
***CASE 2**: $\rho_\epsilon = \rho_x = 0$, HET.	.052 *.045***	.057 *.023***	.300	.037	.059	.051	.162	.100	.118
***CASE 3**: Outlier: $u_t \overset{i.i.d.}{\sim} \begin{cases} N[0,\ 1], \text{ with prob .95} \\ N[0,\ 1000^2], \text{ with prob .05} \end{cases}$	.047 *.044***	.048 *.015***	.088	.043	.083	.039	.056	.008	.009
***CASE 4**: Stationary GARCH(1,1): $u_t = \sigma_t\epsilon_t,$ $\sigma_t^2 = 0.666u_{t-1}^2 + 0.333\sigma_{t-1}^2$ $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1).$	.042 *.040***	.046 *.013***	.040	.005	.005	.004	.080	.046	.046
***CASE 5**: Stochastic Volatility: $u_t = exp(w_t/2)\epsilon_t,$ $w_t = 0.5w_{t-1} + v_t,$ $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, $v_t \sim \chi_2(3)$	.043 *.045***	.041 *.021***	.063	.006	.014	.006	.054	.014	.014
***CASE 6**: DEB: $x_{2t} \overset{i.i.d}{\sim} \mathcal{B}(1, 1/3),$ $x_{3t} \overset{i.i.d}{\sim} \mathcal{N}(0, .01^2)$	.047 *.043***	.049 *.022***	.080	.048	.084	.043	.085	.060	.095
***CASE 6 BIS**: $x_{2t} \overset{i.i.d}{\sim} \mathcal{N}(0,1),$ $x_{3t} \overset{i.i.d}{\sim} \chi_2(1)$ $u_t = x_{3t}\epsilon_t,\ \epsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0,1)$	.044 *.040***	.042 *.018***	.687	.020	.044	.152	.421	.171	.173
***CASE 7**:									

$(x_{2t}, x_{3t}) \sim \mathcal{N}(0, I_2)$, $u_t \overset{i.i.d}{\sim} \mathcal{C}$	.058 *.049***	.059 *.021***	.069	.013	.033	.012	.061	.023	.023
***CASE 11**: Non stationary GARCH(1,1): $u_t = \sigma_t \epsilon_t$, $\sigma_t^2 = 0.8u_{t-1}^2 + 0.8\sigma_{t-1}^2$, $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$.	.054	.057	.003	.000	.001	.000	.060	.016	.016
***CASE 12**: Exp. Variance: $u_t = exp(.2 * t)\epsilon_t$, $\epsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0,1)$	.049	.051	.017	.000	.000	.000	.132	.014	.014

*: cases when mediangale condition holds.
**: sizes using asymptotic critical values ($\chi^2(3)$).

In the previous cases, sizes of tests derived from sign-based finite-sample methods are exactly controlled, whereas asymptotic tests may greatly under reject the null hypothesis. This remark especially holds for cases involving strong heteroskedasticity (cases 4, 6BIS). The asymptotic versions of sign-based tests suffer from the same under rejection than other asymptotic test, suggesting that, for small samples ($T = 50$), the distribution of the test statistic is really far from its asymptotic limit. Hence, the sign-based method that deals directly with this distribution has clearly an advantage on asymptotic methods. When the dependence in the disturbance process is highly nonlinear (Case 6BIS), the BART method based on a kernel estimation of the LAD asymptotic covariance matrix is not reliable anymore.

Autocorrelation in the residuals: when autocorrelation is present in the residuals, the finite-sample sign based method does not lead to exact levels but is still asymptotically valid. We illustrate here the advantages of the finite-sample sign-based inference involving a HAC correction compared to other asymptotically justified methods of inference.

Table 4:: Level of Conditional Tests of $H_0 : \beta = (1, 2, 3)'$. DGP for which sign-based inference is asymptotically valid.

$y_t = x_t\beta + u_t,\ t = 1, \ldots, T.$	*sign*		*LAD*				*OLS*		
$T = 50$	SF	HAC	OS	DMB	MBB5	BART	IID	WH	BART
^{a}**CASE 8**: $\rho_\epsilon = .5,\ \rho_x = 0$, HOM	.126 -	.022 *.019*	.171	.124	.118	.085	.201	.240	212
^{a}**CASE 9**: $\rho_\epsilon = \rho_x = .5$, HET	.218 -	.026 *.017*	.440	.131	.097	.108	.407	.328	.276
^{a}**CASE 10**[6]:	.521	.012	.553	.516	.339	.355	.649	.677	.534

[6]In that case of high persistence, all automatic bandwith parameters are restricted to be less than 10 to avoid inversibility problems.

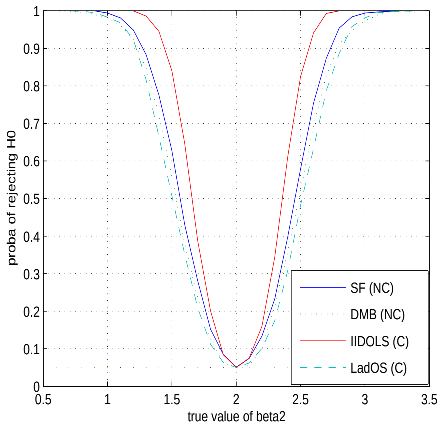
$\rho_\epsilon = .9,\ \rho_x = 0$, HOM	-	*.003*		

^{a}: cases when mediangale condition fails.
$**$: sizes using asymptotic critical values ($\chi^2(3)$).

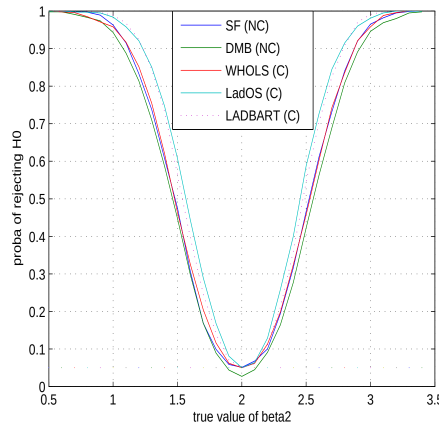
When linear dependence is present in the data, the SF sign-statistic is not adapted anymore and one has to use the $SHAC$ sign-statistic corrected for autocorrelation. The test that yields the best results is the one obtained by using $SHAC$ statistic associated with the finite-sample sign-based method. Whereas all the others Wald tests over reject the null, the latter test seems to control the level avoiding a large under rejection on the contrary to its asymptotic version. We conclude that there really seems to exist important differences between using critical values from the asymptotic distribution of $SHAC$ statistic and critical values derived from the distribution of the $SHAC$ statistic for a fixed number of independent signs. Besides, we underline the dramatic over rejections of asymptotic Wald tests based on HAC estimation of the asymptotic covariance matrix when the data set involves a small number of observations. These results suggest, in a sense, that when the data suffer from both a small number of observations and linear dependence. The first problem to solve is the finite-sample distortion, which is not what is usually done.

7.2. Power

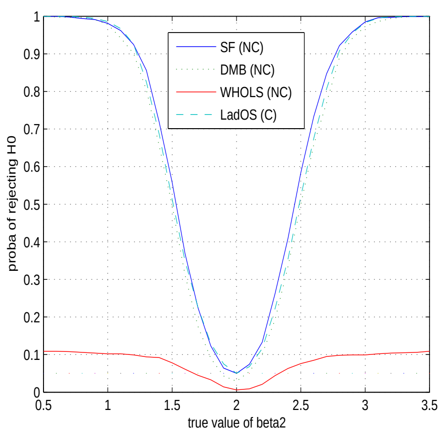
Then, we illustrate the **power** of these tests. We are particularly interested in comparing the sign-based inference to kernel and bootstrap methods. Others methods may not be reliable even in terms of level. We consider the simultaneous hypothesis H_0 as before. The true process is obtained by fixing β_1 and β_3 at the tested value, i.e $\beta_1 = 1$ and $\beta_3 = 3$, and letting vary β_2. Simulated power is given by a graph with β_2 in abscissa. The power functions presented here are locally adjusted for the level when needed. That allows comparisons between methods. However, we have to keep in mind that only the sign-based methods lead to exact confidence levels without adjustment. Other methods may over reject the null hypothesis and do not control the level of the test, or under reject it, and consequently loose power.



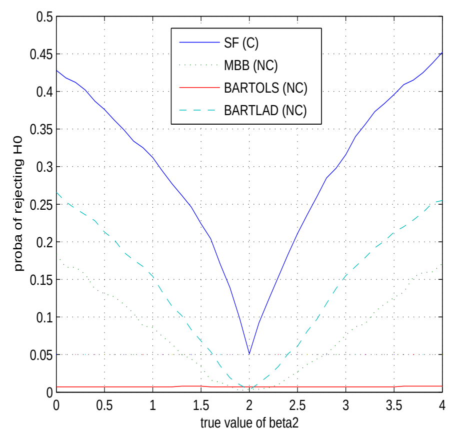
(a) CASE 1 : normal



(b) CASE 2 : normal het

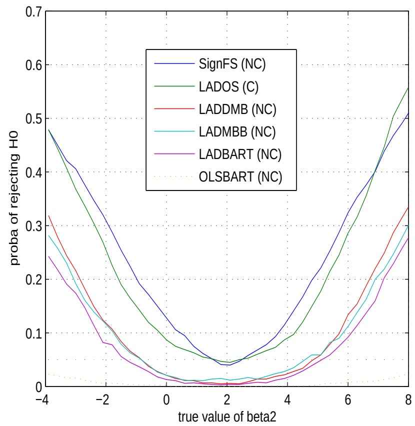


(c) CASE 3 : outliers

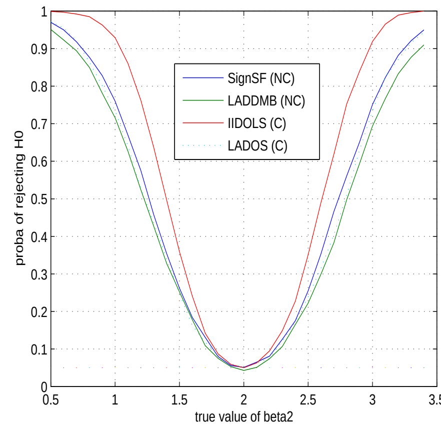


(d) CASE 4 : stationary garch

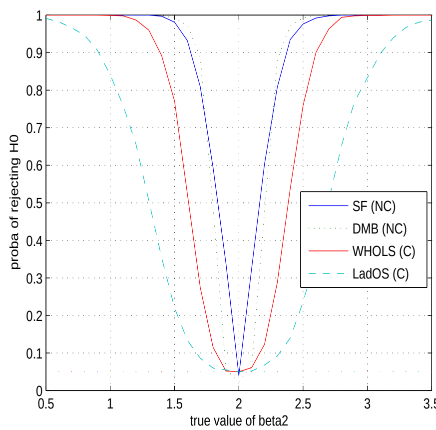
Sign-based inference has a totally comparable power performance with usual methods in cases 1, 2, 3, 8 with the advantage that the level is exactly controlled for any sample size. That can make a great difference in small samples. Remember that any HAC correction has only an asymptotic justification. In very heteroskedastic cases (4, 5, 11, 12), sign-based inference greatly dominates other methods: levels of coverage are exactly controlled and their power function largely exceed others, even with locally adjusted levels. When autocorrelation is present in the data, the sign-based inference based on $D_S[\beta, (X'X)^{-1}]$ looses the exactness of the level. However, when a HAC correction is included in the sign statistic, the method seems to lead to comparable power performance than usual methods such as kernels, moving block bootstrap, along with a better control of the level. Only for very high autocorrelation (close to unit root process), the sign-based inference



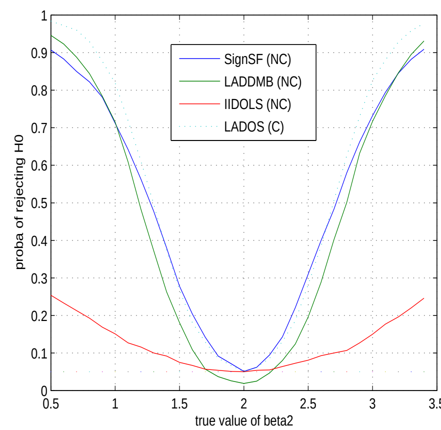
(a) CASE 5 : Stochastic Volatility



(b) CASE 6 : DEB



(c) CASE 6 : BIS

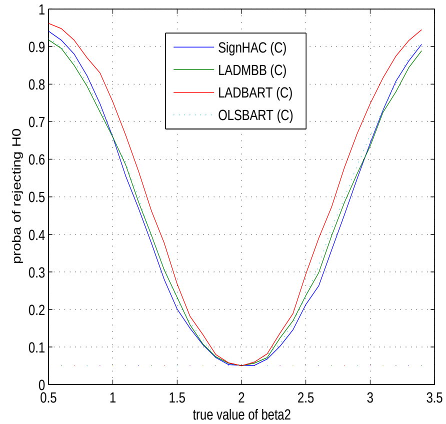


(d) CASE 7 : CAUCHY

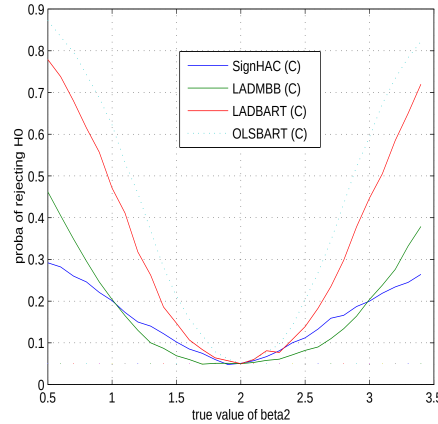
is not adapted anymore.

7.3. Confidence intervals

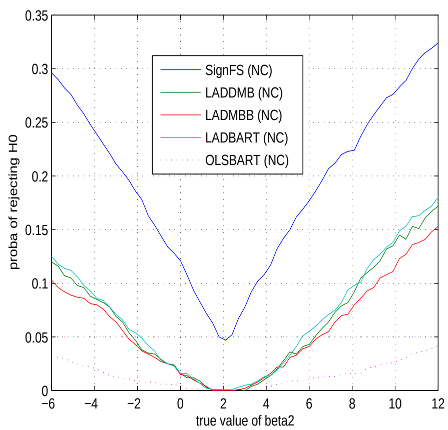
As the sign-based confidence regions are by construction of coverage level higher that $1-\alpha$ whenever inference is exact, a performance indicator for confidence intervals may be the spread of those confidence intervals. Thus, we want to compare the spread of confidence intervals obtained by projecting the sign-based simultaneous confidence regions to those based on t-statistics on the LAD estimator. We use 1000 simulations, and report the means and the empirical standard deviations of those spreads. We only consider the stationary examples. Remark that in the non



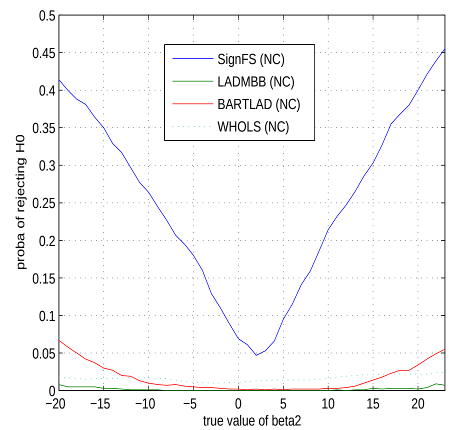
(e) CASE 9: mild autocorrelation



(f) CASE 10 : high autocorelation



(a) CASE 11 : non stationary GARCH



(b) CASE 12 : exponential variance

stationary cases, inference based on t-statistics may not mean anything. In Tables 5, 6, 7, we report average spread of confidence intervals for each β_k and average coverage levels.

Table 5:: Spread of confidence intervals for β_1

$y_t = x_t\beta + u_t,\ t = 1, \ldots, T.$ $T = 50$		*Proj. Based sign*		*t-stat. of LAD Bootstrap*		*t-stat. of LAD Kernel*
		FS	HAC	DMB	MBB	BART
*CASE 1:	aver. spread	1.29	1.16	.81	.79	.82

$\rho_\epsilon = \rho_x = 0$, HOM *(st. dev.)* **cov. lev.**	*(.21)* **1.0**	*(.14)* **1.0**	*(.23)* **.97**	*(.21)* **.95**	*(.15)* **.97**
***CASE 2**: $\rho_\epsilon = \rho_x = 0$, HET	.76 *(.14)* **1.0**	.66 *(.15)* **1.0**	.43 *(.09)* **.98**	.42 *(.10)* **.97**	.50 *(.11)* **.99**
***CASE 3**: Outlier $u_t \overset{i.i.d.}{\sim} \begin{cases} N[0,\ 1], \text{ with prob } .95 \\ N[0,\ 1000^2], \text{ with prob } .05 \end{cases}$	1.26 *(.26)* **1.0**	1.15 *(.25)* **1.0**	.92 *(.80)* **.98**	.88 *(.67)* **.97**	.88 *(.17)* **.97**
***CASE 4**: Stationary GARCH(1,1): $u_t = \sigma_t \epsilon_t$, $\sigma_t^2 = 0.666 u_{t-1}^2 + 0.333 \sigma_{t-1}^2$ $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$.	50.4 *(101.0)* **1.0**	49.5 *(100.2)* **.99**	30.6 *(64.6)* **1.0**	35.0 *(76.7)* **1.0**	29.3 *(70.3)* **1.0**
***CASE 5**: Stochastic Volatility: $u_t = exp(w_t/2)\epsilon_t$, $w_t = 0.5 w_{t-1} + v_t$ $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, $v_t \sim \chi_2(3)$	27.3 *(14.4)* **1.0**	22.8 *(12.2)* **1.0**	13.3 *(6.4)* **.99**	15.1 *(9.6)* **.98**	15.7 *(7.5)* **.99**
***CASE 6**: DEB: $x_{2t} \overset{i.i.d}{\sim} \mathcal{B}(1, 1/3)$, $x_{3t} \overset{i.i.d}{\sim} \mathcal{N}(0, .01^2)$	1.64 *(.29)* **1.0**	1.42 *(.32)* **1.0**	1.01 *(.26)* **.96**	.99 *(.31)* **.94**	1.03 *(.21)* **.96**
***CASE 7**: $(x_{2t}, x_{3t}) \sim \mathcal{N}(0, I_2)$, $u_t \overset{i.i.d}{\sim} \mathcal{C}$	2.20 *(.59)* **1.0**	1.88 *(.56)* **1.0**	1.25 *(.32)* **.98**	1.21 *(.38)* **.97**	1.39 *(.37)* **.99**
^{a}**CASE 8**: $\rho_\epsilon = .5,\ \rho_x = 0$, HOM	1.59 *(.30)* **.99**	1.63 *(.38)* **.99**	.99 *(.25)* **.86**	1.17 *(.34)* **.90**	1.23 *(.36)* **.91**
^{a}**CASE 9**: $\rho_\epsilon = \rho_x = .5$, HET	1.25 *(.31)* **1.0**	1.23 *(.41)* **.98**	.68 *(.17)* **.93**	.79 *(.24)* **.95**	.94 *(.33)* **.97**
^{a}**CASE 10**: $\rho_\epsilon = .9,\ \rho_x = 0$, HOM	2.46 *(.84)* **.68**	3.00 *(1.06)* **.74**	1.52 *(.57)* **.47**	2.46 *(1.00)* **.66**	2.89 *(1.46)* **.71**

Table 6:: Spread of confidence intervals for β_2

$y_t = x_t\beta + u_t,\ t = 1,\dots,T.$ $T = 50$		*Proj. Based sign* FS	HAC	*t-stat. of LAD Bootstrap* DMB	MBB	*t-stat. of LAD Kernel* BART
CASE 1: $\rho_\epsilon = \rho_x = 0$, HOM	aver. spread	1.52	1.36	.90	.88	.88
	(st.dev.)	*(.27)*	*(.28)*	*(.21)*	*(.24)*	*(.19)*
	cov. lev.	**1.0**	**1.0**	**.97**	**.96**	**.96**
CASE 2: $\rho_\epsilon = \rho_x = 0$, HET		1.43	1.26	.94	.90	.92
		(.29)	*(.28)*	*(.24)*	*(.27)*	*(.29)*
		1.0	**1.0**	**.97**	**.95**	**.95**
CASE 3: Outlier $u_t \overset{i.i.d.}{\sim} \begin{cases} N[0,\ 1],\ \text{with prob } .90 \\ N[0,\ 100^2],\ \text{with prob } .10 \end{cases}$		1.37	1.24	.94	.98	.88
		(.31)	*(.29)*	*(.79)*	*(1.36)*	*(.20)*
		1.0	**.99**	**.98**	**.97**	**.979**
CASE 4: Stationary GARCH(1,1): $u_t = \sigma_t\epsilon_t$, $\sigma_t^2 = 0.666u_{t-1}^2 + 0.333\sigma_{t-1}^2$, $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$.		58.5	55.9	33.4	38.3	32.6
		(118)	*(115)*	*(74.6)*	*(82.6)*	*(76.9)*
		1.0	**.99**	**1.0**	**1.0**	**1.0**
CASE 5: Stochastic Volatility: $u_t = exp(w_t/2)\epsilon_t$, $w_t = 0.5w_{t-1} + v_t$, $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, $v_t \sim \chi_2(3)$		30.4	29.4	15.9	20.7	15.4
		(16.7)	*(17.6)*	*(15.9)*	*(28.0)*	*(7.8)*
		.98	**1.0**	**1.0**	**1.0**	**1.0**
CASE 6: DEB: $x_{2t} \overset{i.i.d}{\sim} \mathcal{B}(1, 1/3)$, $x_{3t} \overset{i.i.d}{\sim} \mathcal{N}(0, .01^2)$		2.82	2.48	1.70	1.64	1.68
		(.50)	*(.51)*	*(.36)*	*(.43)*	*(.33)*
		1.0	**1.0**	**.98**	**.96**	**.96**
CASE 7: $(x_{2t}, x_{3t}) \sim \mathcal{N}(0, I_2)$, $u_t \overset{i.i.d}{\sim} \mathcal{C}$		2.75	2.33	1.47	1.41	1.52
		(.82)	*(.78)*	*(.46)*	*(.57)*	*(.49)*
		1.0	**1.0**	**.98**	**.98**	**.98**
^{a}**CASE 8**: $\rho_\epsilon = .5,\ \rho_x = 0$, HOM		1.71	1.47	1.00	.96	.91
		(.32)	*(.31)*	*(.26)*	*(.26)*	*(.23)*
		1.0	**1.0**	**.98**	**.97**	**.96**
^{a}**CASE 9**: $\rho_\epsilon = \rho_x = .5$, HET		1.46	1.64	1.12	1.23	1.11
		(.40)	*(.51)*	*(.33)*	*(.42)*	*(.55)*
		.99		**.88**	**.89**	**.83**
^{a}**CASE 10**: $\rho_\epsilon = .9,\ \rho_x = 0$, HOM		2.42	2.00	1.41	1.56	1.21
		(.82)	*(.68)*	*(.56)*	*(.60)*	*(.47)*

	.99	**1.0**	**.95**	**.97**	**.87**

Table 7:: Spread of confidence intervals for β_3

$y_t = x_t\beta + u_t,\ t = 1,\ldots,T.$ $T = 50$	*Proj. Based* *sign*		*t-stat. of LAD* *Bootstrap*		*t-stat. of LAD* *Kernel*
	FS	HAC	DMB	MBB	BART
***CASE 1**: av. spread	1.40	1.02	.89	.85	.87
$\rho_\epsilon = \rho_x = 0$, HOM *(st. dev.)*	*(.29)*	*(.29)*	*(.22)*	*(.24)*	*(.22)*
cov. lev.	**1.0**	**1.0**	**.97**	**.95**	**.96**
***CASE 2**:	.74	.48	.43	.41	.50
$\rho_\epsilon = \rho_x = 0$, HET	*(.17)*	*(.18)*	*(.11)*	*(.12)*	*(.11)*
	1.0	**1.0**	**.99**	**.97**	**.99**
***CASE 3**: Outlier:	1.05	.91	.98	1.04	.88
$u_t \overset{i.i.d.}{\sim} \begin{cases} N[0,\,1], \text{ with prob } .95 \\ N[0,\,1000^2], \text{ with prob } .05 \end{cases}$	*(.30)*	*(.30)*	*(1.29)*	*(2.73)*	*(.24)*
	.98	**.96**	**.98**	**.97**	**.97**
***CASE 4**: Stationary GARCH(1,1):					
$u_t = \sigma_t\epsilon_t,$	57.3	56.1	25.9	41.5	32.3
$\sigma_t^2 = 0.666u_{t-1}^2 + 0.333\sigma_{t-1}^2$	*(122)*	*(117)*	*(61)*	*(84)*	*(78)*
$\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1).$	**.93**	**.94**	**1.0**	**1.0**	**1.0**
***CASE 5**: Stochastic Volatility:					
$u_t = exp(w_t/2)\epsilon_t,$	33.1	27.0	15.5	19.1	15.6
$w_t = 0.5w_{t-1} + v_t$	*(18.1)*	*(15.8)*	*(10.7)*	*(19.3)*	*(7.5)*
$\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, $v_t \sim \chi_2(3)$	**1.0**	**1.0**	**.99**	**.99**	**.99**
***CASE 6**: DEB:					
$x_{2t} \overset{i.i.d}{\sim} \mathcal{B}(1,1/3),$	188.5	162.9	108.7	104.2	105.67
$x_{3t} \overset{i.i.d}{\sim} \mathcal{N}(0,.01^2)$	*(32.3)*	*(34.4)*	*(25.6)*	*(27.7)*	*(24.5)*
	1.0	**1.0**	**.97**	**.96**	**.96**
***CASE 7**:					
$(x_{2t}, x_{3t}) \sim \mathcal{N}(0, I_2),$	2.59	1.95	1.47	1.42	1.53
$u_t \overset{i.i.d}{\sim} \mathcal{C}$	*(.82)*	*(.74)*	*(.45)*	*(.53)*	*(.47)*
	1.0	**.99**	**.98**	**.97**	**.99**
^{a}**CASE 8**:	1.46	1.05	.94	.86	.81
$\rho_\epsilon = .5,\ \rho_x = 0$, HOM	*(.30)*	*(.28)*	*(.24)*	*(.23)*	*(.21)*
	1.0	**.99**	**.99**	**.97**	**.95**

[a]**CASE 9**: $\rho_\epsilon = \rho_x = .5$, HET	1.56 *(.40)* **1.0**	.99 *(.35)* **.94**	.96 *(.24)* **.98**	.96 *(.26)* **.98**	1.01 *(.36)* **.97**
[a]**CASE 10**: $\rho_\epsilon = .9,\ \rho_x = 0$, HOM	2.69 *(.95)* **1.0**	2.41 *(.96)* **.99**	1.51 *(.61)* **.98**	1.53 *(.63)* **.98**	1.27 *(.61)* **.91**

Spreads of confidence intervals obtained by projection are larger than asymptotic confidence intervals. This is due to the fact that they are by construction conservative confidence intervals. That can be seen as the price to pay to conduct a fully exact inference procedure.

8. Examples

8.1. Regional data set: β convergence across US States

With the neoclassical growth model as theoretical background, Barro and Sala-i Martin (1991) test β convergence between levels of per capita output across 48 US States for different time periods between 1880 and 1988. They use nonlinear least squares to estimate equations of the form

$$(1/T).\ln(y_{i,t}/y_{i,t-T}) = a - [\ln(y_{i,t-T})] \times [(1 - e^{-\beta T})/T](+ \text{ other variables}). \tag{8.3}$$

Their *Basic Regression* does not include other variables. They also consider a specification with regional dummies. The *Basic Regression* model supposes the 48 States share a common per capita level of personal income at steady state while the second specification allows for regional differences in steady state levels. Their regressions involve 48 observations and are run for each 20-year or 10-year period between 1880 and 1988. They tend to accept a positive β and conclude on a convergence between levels of per capita personal income across US States. However, both the NLLS method and tests they perform are only asymptotically justified and can be unreliable for 48 observations. We propose to perform a finite-sample based (conservative) sign test and to see whether their conclusion still holds. We consider the linear equation

$$(1/T).\ln(y_{i,t}/y_{i,t-T}) = a + \gamma[\ln(y_{i,t-T})] \times (+ \text{ other variables}), \tag{8.4}$$

and compute projection-based CI for $\beta = -(1/T)\ln(\gamma T + 1)$ with both specifications. Results are reported in Table 8.1.

Table 8.1: Regressions for Personal income across U.S. States, 1880-1988

Period	*Basic equation*	*Equation with regional dummies*
	95% projection-based CI(β)	
	90% projection-based CI(β)	
1880-1900	[-.00681, .01234]	[-.01230, .02105]
	[-.00631, .01084]	[-.01230, .01811]
1900-1920	[.00917, .03129]	[-.00897, .10633]
	[.00953, .02730]	[-.00524, .05746]
1920-1930	[-.03006, .00180]	[-.04629, .03885]
	[-.02855, -.00037]	[-.04578, .03850]
1930-1940	[.00431, .02340]	[-.01891, .05822]
	[.00441, .02292]	[-.01894, .03513]
1940-1950	[.02914, .06016]	[.00830, .06203]
	[.02926, .05888]	[.00896, .05960]
1950-1960	[.00835, .03518]	[-.00443, .05098]
	[.00835, .03387]	[-.00049, .05002]
1960-1970	[.00987, .03774]	[-.01121, .04310]
	[.01044, .03469]	[-.01089, .04079]
1970-1980	[.00205, .03460]	[-.01307, .07388]
	[.00620, .02881]	[-0.01994, .04079]
1980-1988	[-.05524, .05033]	[-.04723, .13439]
	[-.05402, .04378]	[-.04501, .07646]

The results we find for the *Basic Regression* are very close to those of Barro and Sala-i Martin (1991). We fail to reject $\beta = 0$ at 5%-level, for the 1880-1900, 1920-1930, 1980-1988 periods whereas Barro and Sala-i Martin (1991) fail to reject $\beta = 0$ at 5%(asymptotic)-level for the 1920-1930 and 1980-1988 periods. Our results differ only for the 1880-1900 period. That may be due to the quality of data and the possible presence of outliers affecting least squares methods in that early period of time. Results are less clear when regional dummies are included. We fail to reject $\beta = 0$ at 5%-level 8 times over 9 whereas Barro and Sala-i Martin (1991) fail to reject 5 times over 9. In the best case (for the theory), that may be due to the conservativeness of our tests. However, an objective look on those results may lead to a rejection of the β convergence with possibly different regional steady state levels.

8.2. Huge heteroskedasticity: testing a drift with stochastic volatility in the error term

In that example, we want to test the possible presence of a drift parameter on the Standard and Poor's Composite Price Index (SP), 1928-1987. The data have been used by Gallant, Hsieh and Tauchen (1997) and Valery and Dufour (2004) to fit a stochastic volatility model. So, a huge amount of heteroskedasticty may be suspected. The data set was provided by P. Valéry and 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 only a drift.

$$y_t = a + u_t, \quad t = 1, \ldots, 16127; \tag{8.5}$$

where $\{u_t\}, t = 1, \ldots, 16127$ may present a stochastic volatility. We derive confidence interval for the drift parameter a with the finite-sample sign-based method and we compare them with the ones obtained by Wald techniques applied to LAD and OLS estimates. If the mediangale Assumption **2.5** holds, the sign-based confidence interval exactly control the level. In that large sample, we expect finite-sample sign-based inference to lead to similar results than asymptotic LAD-based inference. The results are reported in Table 8.2.

Methods	*98% Confidence Intervals*
Sign (FS)	[.0232, .0633]
LAD	
with OS covariance matrix estimator	[.0244, .0633]
with DMB covariance matrix estimator	[.0247, .0630]
with kernel covariance matrix estimator	[.0247, .0630]
OLS	[-.0028, .0355]

As expected, results of LAD-based and sign-based tests are very close. They both accept the presence of a positive drift parameter at 98%-level whereas the OLS-based inference cannot reject the absence of a drift.

9. Conclusion

In this paper, we have proposed valid sign-based inference procedures for the β parameter in a linear median regression. We showed that the procedure yields exact tests in finite samples for mediangale processes and remains asymptotically valid for more general processes including stationary ARMA disturbances. We studied the conditions under which consistency and asymptotic normality hold. In particular, we showed that they do require less assumptions on moment existence of the disturbance process than usual LAD asymptotic theory. Simulation studies indicate that the proposed tests and confidence sets are more reliable than usual methods (LS, LAD) even when using the bootstrap. Despite the programming complexity of sign-based methods, we advocate their use when an amount of heteroskedasticity is suspected in the data and the number of available observations is small such as regional data sets. As an example, we reinvestigate whether a β convergence between levels of per capita personal income across US States occurred between 1880 and 1988.

Appendix

A. Appendix : Finite-Sample results

A.1. Proof of Proposition 2.6

We use the fact that, as $\{X\}$ is strictly exogenous, $\{u\}$ does not Granger cause $\{X\}$. It follows directly that $l(s_t|u_{t-1}, \ldots, u_1, x_t, \ldots, x_1) = l(s_t|u_{t-1}, \ldots, u_1, x_n, \ldots, x_1)$ where l stands for the density of $s_t = s(u_t)$.□

A.2. Proof of Proposition 3.2

Consider the vector $[s(u_1), s(u_2), \ldots, s(u_n)]' \equiv (s_1, s_2, \ldots, s_n)'$. By Assumption **2.5**, we have

$$P[u_t > 0|X] = E(P[u_t > 0|u_{t-1}, \ldots, u_1, X]) = 1/2$$

and that

$$P[u_t > 0|u_{t-1}, \ldots, u_1, X] = 1/2 = P[u_t > 0|s_{t-1}, \ldots, s_1, X], \forall t \geq 2.$$

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

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

Then,

$$l(s_1, s_2, \ldots, s_n|X) = \prod_{t=1}^{n} 1/2^{(1-s_t)/2}[1 - 1/2]]^{(1+s_t)/2} = \prod_{t=1}^{n} l(s_t|X).$$

Therefore, $\{s_t,\ t = 1, \ldots, n\}$ are $i.i.d$ (conditional on $X = [x_1, \ldots, x_n]'$) and Proposition **3.2** holds. □

A.3. Proof of Proposition 3.3

Consider model 2.1 with u being a weak conditional mediangale given X. Let show that $[\tilde{s}(u_1), \tilde{s}(u_2), \ldots, \tilde{s}(u_n)]$ can have the same role in Proposition **3.2** as $[s(u_1), s(u_2), \ldots, s(u_n)]$ under Assumption **2.5**. From equation 3.11, we have:

$$\tilde{s}(u_t, V_t) = s(u_t) + [1 - s(u_t)^2]s(V_t - .5).$$

Hence

$$P[\tilde{s}(u_t, V_t) = 1|u_{t-1}, \ldots, u_1, X] = P[s(u_t) + [1 - s(u_t)^2]s(V_t - .5)|u_{t-1}, \ldots, u_1, X].$$

As $(V_1, \ldots, V_n)$ is independent of $(u_1, \ldots, u_n)$ and $\mathcal{U}(0,1)$ distributed, it follows

$$P[\tilde{s}(u_t, V_t) = P[u_t > 0 | u_{t-1}, \ldots, u_1, X] + \frac{1}{2} P[u_t = 0 | u_{t-1}, \ldots, u_1, X]. \tag{A.1}$$

Let $p_t = P[u_t = 0 | u_{t-1}, \ldots, u_1, X]$, the weak conditional mediangale assumption given X yields:

$$P[u_t > 0 | u_{t-1}, \ldots, u_1, X] = P[u_t < 0 | u_{t-1}, \ldots, u_1, X] = \frac{1 - p_t}{2}. \tag{A.2}$$

Reporting A.2 in (A.1) yields

$$P[\tilde{s}(u_t, V_t) = 1 | u_{t-1}, \ldots, u_1, X] = \frac{1 - p_t}{2} + \frac{p_t}{2} = \frac{1}{2}. \tag{A.3}$$

In a similar way,

$$P[\tilde{s}(u_t, V_t) = -1 | u_{t-1}, \ldots, u_1, X] = \frac{1}{2}. \tag{A.4}$$

The remaining of the proof is the same as the proof of Proposition **3.2**.□

A.4. Proof of Proposition 4.1

This proof is similar to Boldin et al. (1997) one in case of $i.i.d$ disturbances and includes the latter as special case. We begin with a single explanatory variable case ($p = 1$) which is more educational and presents the basic ideas. The case with $p > 1$ is just an adaptation of the same ideas to multidimensional notions. Let us consider model (2.1) under the mediangale Assumption **2.5**. In the single parameter case, the locally optimal sign-based test (conditional on X) of $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$ is well defined. Among the tests with a given confidence level α, the power function of the locally optimal sign-based test has the highest slope around 0. The conditional power function of a sign-based test can be written

$$P[s(y) \in W | \beta, X] \,, \tag{A.5}$$

where W is the critical region with level α. Hence, we should include in W the sign vectors for which $\frac{d}{d\beta} P[S(y) = s | 0, X]$ are as large as possible. Under the mediangale Assumption **2.5** and assuming the existence of continuous densities for the disturbances, we have around 0

$$\begin{aligned}
P[S(y) = s | \beta, X] &= \prod_{i=1}^{n} [P(y_i > 0 | \beta, X)]^{(1+s_i)/2} [P(y_i < 0 | \beta, X)]^{(1-s_i)/2} && \text{(A.6)} \\
&= \frac{1}{2^n} \prod_{i=1}^{n} [1 + 2 f_i(0|X) x_i s_i \beta + o(\beta)] && \text{(A.7)} \\
&= \frac{1}{2^n} [1 + 2 \sum_{i=1}^{n} f_i(0|X) x_i s_i \beta + o(\beta)]. && \text{(A.8)}
\end{aligned}$$

And by identification,

$$\frac{d}{d\beta}P[S(y) = s|0, X] = 2^{-n+1}\sum_{i=1}^{n} f_i(0|X)x_i s_i \; . \tag{A.9}$$

Therefore, the required test has the form

$$W = \{s = (s_1, \ldots, s_n)| \sum_{i=1}^{n} f_i(0|X)x_i s_i| > c_\alpha\} \; , \tag{A.10}$$

or equivalently,

$$W = \{s|s'(y)\tilde{X}\tilde{X}'s(y) > c'_\alpha\} \; , \tag{A.11}$$

where c_α and c'_α are defined by the significance level.
Note that when the disturbances have a common conditional density at 0, $f(0|X)$, we find the results of Boldin et al. (1997). The locally optimal sign-based test is given by,

$$W = \{s|s'(y)XX's(y) > c'_\alpha\} \; . \tag{A.12}$$

The statistic does not depend on the conditional density evaluated at zero.
When $p > 1$, we need an extension of the notion of slope around 0 for a multidimensional parameter. Boldin et al. (1997) propose to restrict to the class of locally unbiased tests and to consider the maximal mean curvature. Thus, a locally unbiased sign-based test satisfies,

$$\frac{dP[W|\beta]}{d\beta} \; |_{\beta=0} = 0 \; , \tag{A.13}$$

and the behavior of the power function around 0 is totally defined by the quadratic term of its Taylor expansion

$$\beta' \; \frac{1}{2}(\frac{d^2P[W|\beta]}{d\beta^2}\Big) \beta = \frac{1}{2^n}4 \sum_{1\leq i \neq} \sum_{j \leq n} [f_i(0|X)s_i\beta' x_i][f_j(0|X)s_j x'_j\beta]. \tag{A.14}$$

The locally most powerful sign-based test in the sense of the mean curvature maximizes the mean curvature which is proportional to the trace of $\frac{d^2P[W|\beta]}{d\beta^2}$ at $\beta = 0$ [see Boldin et al. (1997)]. Taking the trace in expression (A.14), we find (after some computations) it is proportional to

$$\sum_{1\leq i \neq} \sum_{j \leq n} f_i(0|X)f_j(0|X)s_i s_j \sum_{k=1}^{p} x_{ik}x_{jk}. \tag{A.15}$$

By adding the independent of s quantity $\sum_{i=1}^{n}\sum_{k=1}^{p} x_{ik}^2$ to (A.15), we find

$$\sum_{k=1}^{p}\left(\sum_{i=1}^{n} x_{ik} f_i(0|X) s_i\right)^2 = s'(y)\tilde{X}\tilde{X}'s(y). \tag{A.16}$$

Hence, the locally optimal sign-biased test in the sense developed by Boldin et al. (1997) for heteroskedastic signs, is

$$W = \{s : s'(y)\tilde{X}\tilde{X}'s(y) > c'_\alpha\} \,. \tag{A.17}$$

The same reasoning with another definition of curvature leads to

$$W = \{s : s'(y)\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'s(y) > c'_\alpha\} \,. \tag{A.18}$$

B. Appendix: Asymptotic results

B.1. Proof of Theorem 6.6

This proof follows the usual steps of an asymptotic normality result for mixing processes [see White (2001)]. In the following, s_t stands for $s(u_t)$.

1. Consider model (2.1). Under Assumption **6.4**, $V_n^{-1/2}$ exists for any n. Let us denote $\mathcal{Z}_{nt} = \lambda' V_n^{-1/2} x'_t s(u_t)$, for some $\lambda \in \mathbb{R}^p$ such that $\lambda'\lambda = 1$. The mixing property of (x'_t, u_t) (condition **6.1**) is reported on $\mathcal{Z}_{nt}$ [Theorem 3.49, in White (2001)]. Hence, $\lambda' V_n^{-1/2} s(u_t) \otimes x_t$ is α-mixing of size $-r/(r-2),\ r > 2$.

2. Condition **6.2** implies

$$E[\lambda' V_n^{-1/2} x'_t s(u_t)] = 0,\ \forall t = 1, \ldots, n,\ \forall n \in \mathbb{N}. \tag{B.19}$$

   Condition **6.3** implies

$$E|\lambda' V_n^{-1/2} x'_t s(u_t)|^r < \Delta < \infty,\ \forall t = 1, \ldots, n,\ \forall n \in \mathbb{N}. \tag{B.20}$$

   Remark that

$$\mathrm{var}(\frac{1}{\sqrt{\mathrm{n}}}\sum_{\mathrm{t}=1}^{\mathrm{n}} \mathcal{Z}_{\mathrm{nt}}) = \mathrm{var}[\frac{1}{\sqrt{\mathrm{n}}}\sum_{\mathrm{t}=1}^{\mathrm{n}} \lambda' \mathrm{V}_\mathrm{n}^{-1/2} \mathrm{s}(\mathrm{u_t}) \otimes \mathrm{x_t}] = \lambda' \mathrm{V}_\mathrm{n}^{-1/2} \mathrm{V}_\mathrm{n} \mathrm{V}_\mathrm{n}^{-1/2} \lambda = 1. \tag{B.21}$$

3. The mixing property of $\mathcal{Z}_{nt}$ and equations (B.19) to (B.21) allow us to apply a Wooldridge-White central limit theorem [Th 5.20 in White (2001)] that yields

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \lambda' V_n^{-1/2} s(u_t) \otimes x_t \to \mathcal{N}(0, 1). \tag{B.22}$$

4. Since λ can be arbitrary chosen provided that $\lambda'\lambda = 1$, Cramér-Wold device implies

$$V_n^{-1/2} n^{-1/2} \sum_{t=1}^{n} s(u_t) \otimes x_t \to \mathcal{N}(0, I_p). \tag{B.23}$$

5. Finally, condition **6.5** states that Ω_n is a consistent estimate of V_n^{-1}. Hence,

$$n^{-1/2} \Omega_n^{1/2} \sum_{t=1}^{n} s(u_t) \otimes x_t \to \mathcal{N}(0, I_p). \tag{B.24}$$

   Under H_0, $s_t = s(y_t - x_t'\beta_0)$. Finally,

$$n^{-1} s'(y - X\beta_0) X \Omega_n X' s(y - X\beta_0) \to \chi_2(p). \tag{B.25}$$

B.2. Proof of Corollary 6.7

Let $\mathcal{F}_t = \sigma(y_0, \ldots, y_t, x_0', \ldots, x_t')$. When the mediangale Assumption **2.5** holds, $\{s(u_t) \otimes x_t, \mathcal{F}_t,\ t = 1, \ldots, n\}$ result from a martingale difference with respect to $\mathcal{F}_t$. Hence,

$$V_n = \mathrm{Var}[\frac{1}{\sqrt{n}} s \otimes X] = \frac{1}{n} \sum_{t=1}^{n} E(x_t s_t s_t' x_t') = \frac{1}{n} \sum_{t=1}^{n} E(x_t x_t') = \frac{1}{n} E(X'X),$$

and $X'X/n$ is a consistent estimate of $E(X'X/n)$. Theorem **6.6** gives $SF(\beta_0) \to \chi_2(p)$. □

B.3. Proof of Theorem 6.8

1. First, we show lemma **B.1** that will be needed in the proof of Theorem **6.8**.

   Lemma B.1 *Let $(F_n)_{n\in\mathbb{N}}$ and F be right continuous distribution functions such that $F_n(-\infty) = F(-\infty)$ and $F_n(+\infty) = F(+\infty)$. Suppose that,*

$$F_n(x) \to_{n\to\infty} F(x),\ \forall x \in \mathbb{R}.$$

   Then, (F_n) converges uniformly to F in $\mathbb{R}$, i.e.

$$sup_{-\infty<x<+\infty} |F_n(x) - F(x)| \to_{n\to\infty} 0.$$

   Proof: Similar proofs can be found in Chung (2001). Suppose on the contrary that there exist $\epsilon > 0$, a sequence $\{n_k,\ k \in \mathbb{N}\}$ of integers tending to $+\infty$, and a real sequence $\{x_k,\ k \in \mathbb{N}\}$, such that for all k:

$$|F_{n_k}(x_k) - F(x_k)| \geq \epsilon > 0 \tag{B.26}$$

If $\{x_k\}$ is not a convergent sequence, consider instead a convergent subsequence. This can be done as $\mathbb{R} \cup \{-\infty, +\infty\}$ is compact. Cases when $x_k \to \infty$ can be excluded as $F_n(+\infty) = F(+\infty)$ and $F_n(-\infty) = F(-\infty)$. Hence, without loss of generality, we can choose $\{x_k\} \to \xi$ where $-\infty < \xi < +\infty$.

Let us consider two sequences $\{r_m^a\}$ and $\{r_m^b\}$ tending to ξ and such that $r_m^a < \xi < r_m^b$. For sufficiently large k, we face the following cases,

- Case 1: $\{x_k\}$ is increasing and $x_k < \xi$:

$$\begin{aligned}\epsilon &\leq F_{n_k}(x_k) - F(x_k) \leq F_{n_k}(\xi^-) - F(r_m^a) \\ &\leq F_{n_k}(\xi^-) - F_{n_k}(\xi) + F_{n_k}(r_m^b) - F(r_m^b) - F(r_m^a).\end{aligned}$$

- Case 2: $\{x_k\}$ is increasing and $x_k < \xi$:

$$\begin{aligned}\epsilon &\leq F(x_k) - F_{n_k}(x_k) \leq F(\xi^-) - F_{n_k}(r_m^a) \\ &\leq F(\xi^-) - F(r_m^a) + F(r_m^a) - F_{n_k}(r_m^a).\end{aligned}$$

- Case 3: $\{x_k\}$ is decreasing and $x_k \geq \xi$:

$$\begin{aligned}\epsilon &\leq F(x_k) - F_{n_k}(x_k) \leq F(r_m^b) - F_{n_k}(\xi) \\ &\leq F(r_m^b) - F(r_m^a) + F(r_m^a) - F_{n_k}(r_m^a) + F_{n_k}(\xi^-) - F_{n_k}(\xi).\end{aligned}$$

- Case 4: $\{x_k\}$ is decreasing and $x_k \leq \xi$:

$$\begin{aligned}\epsilon &\leq F_{n_k}(x_k) - F(x_k) \leq F_{n_k}(r_m^b) - F(\xi) \\ &\leq F_{n_k}(r_m^b) - F_{n_k}(r_m^a) + Fn_k(r_m^a) - F(r_m^a) + F(r_m^a) - F(\xi).\end{aligned}$$

In each case, for fixed k, m can be chosen such that r_m^b and r_m^a are arbitrarily close to ξ. Then, using right continuity properties of F and F_n, the right hand member of each chain of inequalities does not exceed a quantity that tends to zero as $n_k \to \infty$. Thus a contradiction is obtained. We conclude on the uniform convergence of F_n towards F.

2. Let us return to the proof of the theorem. $\tilde{G}_n$ can be rewritten as

$$\begin{aligned}\tilde{G}_n(c_nS_0^n|X_n) = \quad & [\tilde{G}_n(c_nS_0^n|X_n(\omega)) - G(c_nS_0^n)] \\ + \quad & [G(c_nS_0^n) - G_n(c_nS_0^n|X_n(\omega)) \\ + \quad & G_n(c_nS_0^n|X_n).\end{aligned}$$

Since $G(-\infty) = \tilde{G}_n(-\infty) = 0$, $G(+\infty) = \tilde{G}_n(+\infty) = 1$, and $\tilde{G}_n(x|X_n(\omega)) \to G(x)$ $a.e.$, Lemma **B.1** entails that the convergence is uniform. Hence

$$[G(c_nS_0^n) - \tilde{G}_n(c_nS_0^n|X_n)] = o_p(1).$$

The same holds for G_n,

$$[G(c_nS_0^n) - G_n(c_nS_0^n|X_n)] = o_p(1).$$

Hence

$$G_n(c_nS_0^n|X_n) = \tilde{G}_n(c_nS_0^n|X_n) + o_p(1). \tag{B.27}$$

Note that $c_nS_0^n$ is a discrete positive random variable and G_n, its survival function is also discrete. It directly follows from properties of survival functions, that for each $\alpha \in Im(G_n(\mathbb{R}^+))$, i.e. for each point of the image set, we have

$$P[G_n(c_nS_0^n) \leq \alpha] = \alpha. \tag{B.28}$$

Consider now the case when $\alpha \in (0,1) \backslash Im(G_n(\mathbb{R}^+))$. α must be between the two values of a jump of the function G_n. Since G_n is bounded and decreasing, there exist $\alpha_1,\ \alpha_2 \in Im(G_n(\mathbb{R}^+))$, such that

$$\begin{gathered}\alpha_1 < \alpha < \alpha_2, \\ P[G_n(c_nS_0^n) \leq \alpha_1] \leq P[G_n(c_nS_0^n) \leq \alpha] \leq P[G_n(c_nS_0^n) \leq \alpha_2].\end{gathered}$$

More precisely, the first inequality is an equality. Indeed,

$$\begin{aligned}P[G_n(c_nS_0^n) \leq \alpha] = \quad & P[\{G_n(c_nS_0^n) \leq \alpha_1\} \cup \{\alpha_1 < G_n(c_nS_0^n) \leq \alpha\}] \\ = \quad & P[G_n(c_nS_0^n) \leq \alpha_1] + 0,\end{aligned}$$

as $\{\alpha_1 < G_n(c_nS_0^n) \leq \alpha\}$ is a zero-probability event. Applying (B.28) to α_1,

$$P[G_n(c_nS_0^n) \leq \alpha] = P[G_n(c_nS_0^n) \leq \alpha_1] = \alpha_1 \leq \alpha \tag{B.29}$$

Hence, for $\alpha \in (0,1)$, we have

$$P[G_n(c_nS_0^n) \leq \alpha] \leq \alpha. \tag{B.30}$$

Equation (B.30) combined with equation (B.27) allows us to write,

$$P[\tilde{G}_n(c_n S_0^n) \leq \alpha] = P[G_n(c_n S_0^n) \leq \alpha] + o_p(1) \leq \alpha + o_p(1), \tag{B.31}$$

that is,

$$\lim_{n\to\infty} P[\tilde{G}_n(c_n S_0^n) \leq \alpha] \leq \alpha, \tag{B.32}$$

which was to be proved.

B.4. Proof of Theorem 6.9

Let $(U_0, U_1, \ldots, U_N)$ be a vector of $N+1$ $i.i.d$ random variables drawn from a $\mathcal{U}[0,1]$ distribution, S_0^n the observed statistic and $S^n(N) = (S_1^n, \ldots, S_N^n)$, a vector of N independent replicates drawn from $\tilde{G}_n(x)$. The randomized empirical survival function of S_0^n conditional on X, under the null, is given by

$$\tilde{G}^r_{N,n}[x, n, U_0, S_0^n, S^n(N), U(N)|X_n] = 1 - \frac{1}{N}\sum_{j=1}^{N} s(x - c_n S_j^n) + \frac{1}{N}\sum_{j=1}^{N} \delta(c_n S_j^n - x) s(U_j - U_0) \tag{B.33}$$

with $u(x) = 1_{[0,\infty)}(x)$, $\delta(x) = 1_{\{0\}}$. The corresponding randomized empirical $p-$value is

$$\tilde{p}^r_N(x) = \frac{N\tilde{G}^r_{N,n}(x) + 1}{N+1}. \tag{B.34}$$

Usually, validity of Monte Carlo testing is based on the fact the vector $(c_n S_0^n, \ldots, c_n S_N^n)$ is exchangeable. Indeed, in that case, the distribution of ranks is fully specified and yields the validity of empirical $p-$value [see Dufour (2002)]. In our case, it is clear that $(c_n S_0^n, \ldots, c_n S_N^n)$ is not exchangeable and hence, Monte Carlo validity cannot be directly applied. Nevertheless, we will show that asymptotic exchangeability still holds, which will enable us to conclude. To obtain that the vector $(c_n S_0^n, \ldots, c_n S_N^n)$ is asymptotically exchangeable, we show that for any permutation $\pi : [1, N] \to [1, N]$,

$$\lim_{n\to\infty} P[S_0^n \geq t_0, S_1^n \geq t_1, \ldots, S_N^n \geq t_N] - P[S_{\pi(0)}^n \geq t_0, S_{\pi(1)}^n \geq t_1, \ldots, S_{\pi(N)}^n \geq t_N] = 0.$$

First, let rewrite

$$P[S_0^n \geq t_0, S_1^n \geq t_1, \ldots, S_N^n \geq t_N] = E_{X_n}\{P[S_0^n \geq t_0, S_1^n \geq t_1, \ldots, S_N^n \geq t_N, X_n = x_n]\}.$$

Hence, if we use the conditional independence of the signs vectors (replicated and observed), we obtain

$$P[S_0^n \geq t_0, S_1^n \geq t_1, \ldots, S_N^n \geq t_N, X_n = x_n] = \quad P[X_n = x_n]\prod_{i=0}^{N} P[S_i^n \geq t_i | X_n = x_n]$$

$$
\begin{aligned}
&= \quad G_n(t_0|X_n = x_n)\prod_{i=1}^{N}\tilde{G}_n(t_i|X_n = x_n) \\
&\to \quad \prod_{i=0}^{N} G(t_i)\text{with probability one.} \qquad \text{(B.35)}
\end{aligned}
$$

As each survival function converges with probability one to $G(x)$, we finally obtain (B.35). Hence, it is straightforward to see that for $\pi : [1, N] \to [1, N]$, we asymptotically (on n) also have

$$P[S_0^n \geq t_{\pi(0)}, S_{\pi(1)}^n \geq t_1, \ldots, S_{\pi(N)}^n \geq t_N, X_n = x_n] \to \prod_{i=0}^{N} G(t_i)\text{with probability one.}$$

Note that as $G(t)$ is not a function of the realization $X(\omega)$, B.35 is stated unconditional.

$$\lim_{n\to\infty} P[S_0^n \geq t_0, S_1^n \geq t_1, \ldots, S_N^n \geq t_N] - P[S_{\pi(0)}^n \geq t_0, S_{\pi(1)}^n \geq t_1, \ldots, S_{\pi(N)}^n \geq t_N] = 0.$$

Hence, we can apply an asymptotic version of Proposition 2.2.2 in Dufour (2002) that validates Monte Carlo testing for general possibly non continuous statistics. The proof of this asymptotic version follows exactly the same steps as the proofs of Lemma 2.2.1 and Proposition 2.2.2 of Dufour (2002). We just have to replace the exact distributions of randomized ranks, the empirical survival functions and the empirical $p-$values by their asymptotic counterparts and this is sufficient to conclude. Suppose that N, the number of replicates is such that $\alpha(N+1)$, is an integer. Then,

$$\lim_{n\to\infty} \tilde{p}_N^r(c_n S_0^n) \leq \alpha,$$

which was to be proved.

C. Appendix: more simulation results

C.1. Conditional levels for $T = 200$ with $S = 5000$ and $R = 2999$

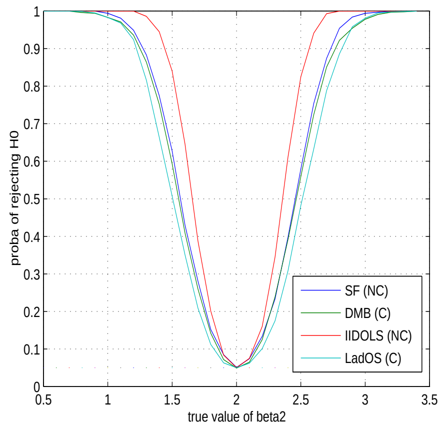
Table 8:: Level of Conditional Tests of $H_0 : \beta = (1, 2, 3)'$.

$y_t = x_t\beta + u_t,\ t = 1, \ldots, T.$ $T = 200$	*sign* SF	*sign* HAC	*LAD* OS	*LAD* DMB	*LAD* MBB5	*LAD* BART	*OLS* IID	*OLS* WH	*OLS* BART
***CASE 1**: $\rho_\epsilon = \rho_x = 0$, HOM.	.049 *.048***	.049 *.042***	.087	.060	.069	.039	.053	.060	.063
***CASE 2**: $\rho_\epsilon = \rho_x = 0$, HET.	.052 *.047***	.058 *.040***	.372	.046	.059	.031	.176	.069	.071
***CASE 3**: Outlier: $u_t \overset{i.i.d.}{\sim} \begin{cases} N[0,\ 1],\ \text{with prob .95} \\ N[0,\ 1000^2],\ \text{with prob .05} \end{cases}$	.053 *.050***	.053 *.043***	.078	.060	.070	.033	.041	.003	.003
***CASE 4**: Stationary GARCH(1,1): $u_t = \sigma_t\epsilon_t,$ $\sigma_t^2 = 0.666u_{t-1}^2 + 0.333\sigma_{t-1}^2$ $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 1).$	.044 *.048***	.045 *.040***	.003	.000	.000	.000	.066	.016	.014
***CASE 5**: Stochastic Volatility: $u_t = exp(w_t/2)\epsilon_t,$ $w_t = 0.5w_{t-1} + v_t,$ $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$, $v_t \sim \chi_2(3)$	.040 *.046***	.040 *.039***	.067	.019	.024	.002	.054	.009	.009
***CASE 6**: DEB: $x_{2t} \overset{i.i.d}{\sim} \mathcal{B}(1, 1/3),$ $x_{3t} \overset{i.i.d}{\sim} \mathcal{N}(0, .01^2)$	.040 *.046***	.040 *.037***	.081	.055	.067	.034	.053	.062	.064
***CASE 6 BIS**: $x_{2t} \overset{i.i.d}{\sim} \mathcal{N}(0, 1),$ $x_{3t} \overset{i.i.d}{\sim} \chi_2(1)$ $u_t = x_{3t}\epsilon_t,\ \epsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0, 1)$	.049 *.050*	.049 *.045*	.924	.016	.020	.051	.293	.070	.070
***CASE 7**: $(x_{2t}, x_{3t}) \sim \mathcal{N}(0, I_2),$ $u_t \overset{i.i.d}{\sim} \mathcal{C}$	.040 *.048***	.044 *.039***	.073	.038	.049	.008	.047	.016	.016
^{a}**CASE 8**: $\rho_\epsilon = .5,\ \rho_x = 0$, HOM	.156 -	.061 *.026***	.193	.160	.098	.069	.210	.219	.116
^{a}**CASE 9**: $\rho_\epsilon = \rho_x = .5$, HET	.186 -	.067 *.024***	.589	.166	.082	.055	.425	.246	.143
^{a}**CASE 10**: $\rho_\epsilon = .9,\ \rho_x = 0$, HOM	.516 -	.039 *.000***	.542	.515	.275	.186	.599	.603	.301

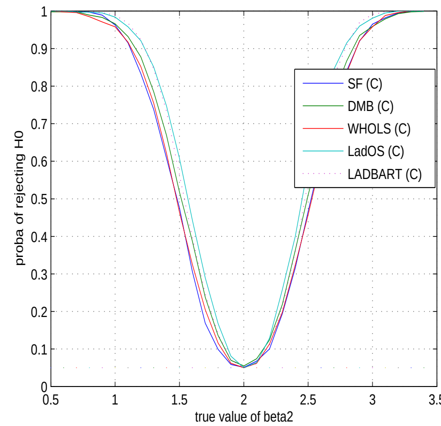
*****CASE 11**: Non stationary GARCH(1,1): $u_t = \sigma_t \epsilon_t$, $\sigma_t^2 = 0.8u_{t-1}^2 + 0.8\sigma_{t-1}^2$, $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$.	.059	.058	.016	.000	.000	.000	.029	.028	.028
***CASE 12**: Exp. Variance: $u_t = exp(.2 * t)\epsilon_t$, $\epsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0,1)$	.056	.056	.028	.000	.000	.000	.013	.019	.019

C.2. Power functions with corrected levels

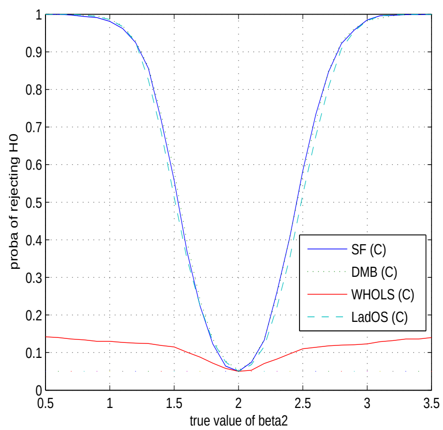
In the following graphics, all the methods are locally corrected in order to exactly fix the level to 5%.



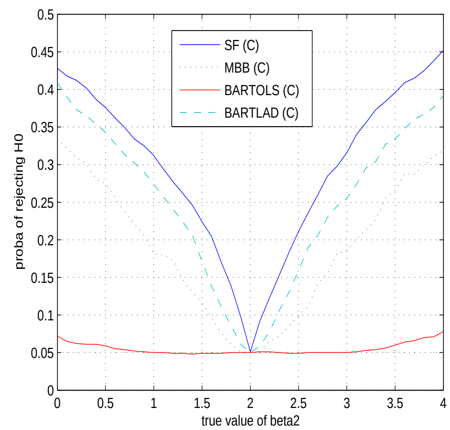
(c) CASE 1 : normal



(d) CASE 2 : normal het



(e) CASE 3 : outliers



(f) CASE 4 : stationary garch

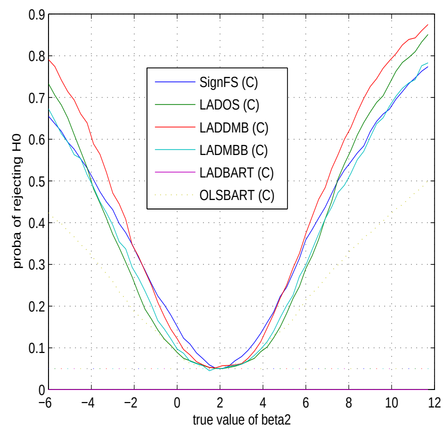
References

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* **LXXX**, 520–534.

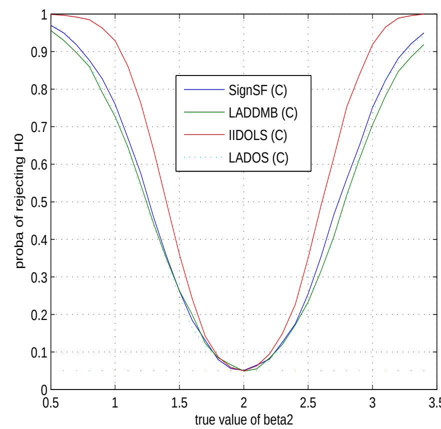
Amemiya, T. (1982), 'Two stage least absolute deviations estimator', *Econometrica* **50**, 689–711.

Andrews, D. (1991), 'Heteroskedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica* **59**, 817–858.

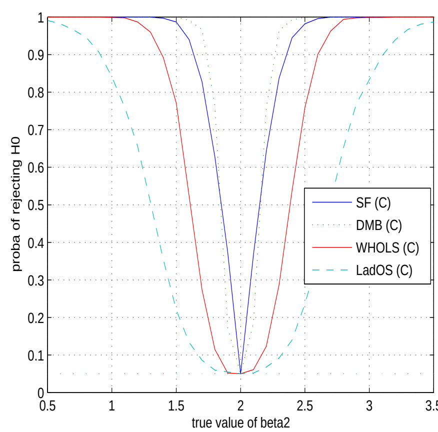
Bahadur, R. and Savage, L. (1956), 'The nonexistence of certain statistical procedures in nonparametric problem', *Annals of Mathematical Statistics* **27**, 1115–1122.



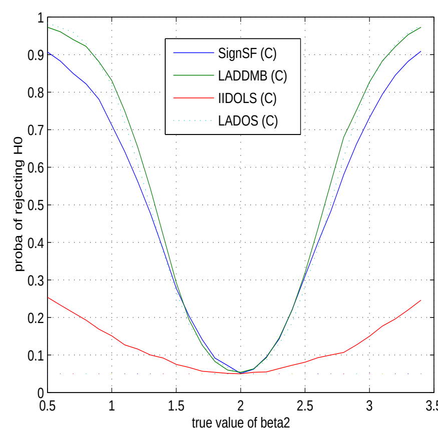
(a) CASE 5 : Stochastic Volatility



(b) CASE 6 : DEB



(c) CASE 6 : BIS



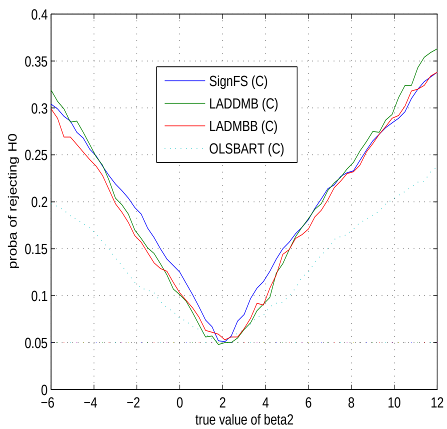
(d) CASE 7 : CAUCHY

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**, 294.

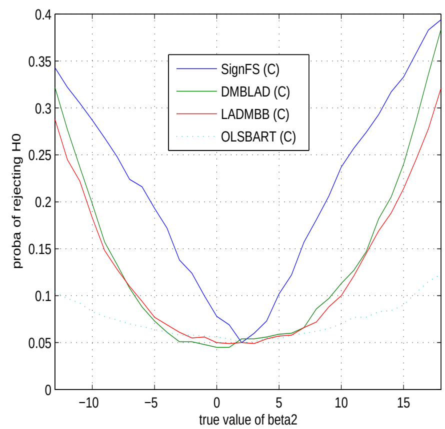
Barro, R. and Sala-i Martin, X. (1991), ‘Convergence across states and regions’, *Brookings Papers on Economic Activity* **1**.

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

Buchinsky, M. (1994), ‘Changes in the us wages stucture 1963-1987: exploration by quantiles regressions’, *Econometrica* **13**, 105–158.



(e) CASE 11 : non stationary GARCH



(f) CASE 12 : exponential variance

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

Buchinsky, M. (2003), 'Recent advances in quantile regression models: A practical guideline for empirical research', *forthcoming, Journal of Human Resources* .

Buchinsky, M. . H. J. (1998), 'An alternative estimator for the censored quantile regression model', *Econometrica* **66**(3), 653–671.

Campbell, B. and Dufour, J.-M. (1991), 'Over-rejections in rational expectations models: A nonparametric approach to the Mankiw-Shapiro problem', *Economics Letters* **35**, 285–290.

Campbell, B. and Dufour, J.-M. (1995), 'Exact nonparametric orthogonality and random walk tests', *Review of Economics and Statistics* **77**, 1–16.

Campbell, B. and Dufour, J.-M. (1997), 'Exact nonparametric tests of orthogonality and random walk in the presence of a drift parameter', *International Economic Review* **38**, 151–173.

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.

Chung, K. (2001), *A course in probability theory*, Academic Press, New-York.

Coakley, C. W. and Heise, M. A. (1996), 'Versions of the sign test in the presence of ties', *Biometrics* **52**(4), 1242–1251.

Coudin, E. and Dufour, J. (2005), Hodges-lehmann robust sign-based estimators in linear median regressions under heteroskedasticity and nonlinear dependence of unknown form, Technical report, mimeo.

De Angelis, D., Hall, P. and Young, G. A. (1993), 'Analytical and bootstrap approximations to estimator distributions in $l1$ regression', *Journal of the American Statistical Association* **88**(424), 1310–1316.

Dielman, T. and Pfaffenberger, R. (1988*a*), 'Bootstrapping in least absolute value regression: An application to hypothesis testing', *Communications in Statistics, Simulation and Computation* **17**, 843–856.

Dielman, T. and Pfaffenberger, R. (1988*b*), 'Least absolute value regression: Necessary sample sizes to use normal theory inference procedures', *Decision Sciences* **19**, 734–743.

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. (1981), 'Rank tests for serial dependence', *Journal of Time Series Analysis* **2**, 117–128.

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

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

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

Dufour, J.-M. (2002), 'Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in econometrics', *Journal of Econometrics* **forthcoming**.

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**, 815–843.

Dufour, J. and Taamouti, M. (2000), Projection-based statistical inference in linear structural models with possibly weak instruments". CIRANO and CIREQ, discussion paper, revised 2001.

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

Engle, R. and Manganelli, S. (2000), CAViaR: Conditional autoregressive value at risk by regression quantiles, *in* 'Econometric Society World Congress 2000 Contributed Papers'.

Fitzenberger, B. (1997*a*), A guide to censored quantile regressions, *in* G. S. Maddala and C. R. Rao, eds, 'Handbook of Statistics 15: Robust Inference', North-Holland, Amsterdam, pp. 405–437.

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

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

Godambe, V. (2001), Estimation of median: quasilikelihood and estimating functions, Technical report, 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.

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

Horowitz, J. (1998), 'Bootstrap methods for median regression models', *Econometrica* **66**, 1327–1352.

Jung, S. (1996), 'Quasi-likelihood for median regression models', *Journal of the American Statistical Association* **91**, 251–257.

Kim, T. and White, H. (2002), Estimation, inference, and specification testing for possibly misspecified quantile regression. Working Paper.

Koenker, R. and Bassett, G. (1982), 'Tests of linear hypotheses and L_1 estimation', *Econometrica* **50**, 1577–1584.

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

Koenker, R. and Hallock, K. (2000), 'Quantile regression: an introduction', *Presented at the Journal of Economic Perspectives Symposium on Econometics Tools.* .

Komunjer, I. (2003), Efficient conditional quantile estimation and specification testing. Working Paper.

Lehmann, E. L. and Stein, C. (1949), 'On the theory of some non-parametric hypotheses', *Annals of Mathematical Statistics* **20**, 28–45.

Meghir, C. and Pistaferri, L. (2004), 'Income variance dynamics and heterogeneity', *Econometrica* .

Newey, W. and West, K. (1987), 'A simple positive semi-definite, heteroskadasticity and autocorrelation consistent covariance matrix', *Econometrica* **55**, 703–708.

Parzen, E. (1957), 'On consistent estimates of the spectrum of a stationary times series', *Annals of Mathematical Statistics* **28**, 329–348.

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

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

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

Powell, J. (1994), *Estimation of Semiparametric Models*, Vol. 4, North-Holland, Amsterdam, the Netherlands, chapter 41.

Pratt, J. W. and Gibbons, J. D. (1981), *Concepts of Nonparametric Theory*, Springer-Verlag, New York.

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

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

Valery, P. and Dufour, J. (2004), Finite-sample and simulation-based tests for stochastic volatility model, Technical report, CIRANO-CIREQ.

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

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

White, H. (1980), 'A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity', *Econometrica* **48**, 817–838.

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

White, H., ed. (1984), *Asymptotic Theory For Econometricians*, Academic Press, New York. Revised Version (2001).

Wright, J. H. (2000), 'Alternative variance-ratio tests using ranks and signs', *Journal of Business and Economic Statistics* **18**(1), 1–9. available at http://ideas.repec.org/a/bes/jnlbes/v18y2000i1p1-9.html.

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