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: January 2004  
Compiled: February 9, 2005, 1:39pm

*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

1. Introduction 1

2. Framework 4
   2.1. Model ................................................................. 4
   2.2. Special cases ......................................................... 7

3. Exact finite-sample sign-based inference 8
   3.1. Motivation .......................................................... 8
   3.2. Distribution-free pivotal functions and nonparametric tests .......... 9

4. Regression sign-based tests 11
   4.1. Regression sign statistics ......................................... 11
   4.2. Monte Carlo tests .................................................... 13

5. Regression sign-based confidence sets and confidence distributions 15
   5.1. Confidence sets and conservative confidence intervals .................. 15
   5.2. Confidence distribution ............................................. 16
   5.3. Numerical illustration .............................................. 17

6. Asymptotic theory 19
   6.1. Asymptotic distributions of test statistics .......................... 20
   6.2. Asymptotic validity of Monte Carlo tests .......................... 21
       6.2.1. Generalities ..................................................... 21
       6.2.2. Asymptotic validity of sign-based inference .................. 22

7. Simulations and comparisons with other methods 23
   7.1. Inference ............................................................. 25
   7.2. Confidence intervals ............................................... 28

8. Conclusion 29

A. Appendix: Finite-Sample results 31
   A.1. Proof of Proposition 2.6 .......................................... 31
   A.2. Proof of Proposition 3.2 .......................................... 31
   A.3. Proof of Proposition 3.3 .......................................... 31
   A.4. Proof of Proposition 4.1 .......................................... 32

B. Appendix: Asymptotic results 34
   B.1. Proof of Theorem 6.6 .............................................. 34
   B.2. Proof of Corollary 6.7 ............................................. 35
   B.3. Proof of Theorem 6.8 .............................................. 35
List of Propositions and Theorems

<table>
<thead>
<tr>
<th>Number</th>
<th>Type</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Definition</td>
<td>Martingale difference</td>
<td>4</td>
</tr>
<tr>
<td>2.2</td>
<td>Definition</td>
<td>Strict mediangale</td>
<td>4</td>
</tr>
<tr>
<td>2.3</td>
<td>Definition</td>
<td>Strict conditional mediangale</td>
<td>5</td>
</tr>
<tr>
<td>2.4</td>
<td>Definition</td>
<td>Weak conditional mediangale</td>
<td>5</td>
</tr>
<tr>
<td>2.5</td>
<td>Assumption</td>
<td>Conditional strict mediangale</td>
<td>6</td>
</tr>
<tr>
<td>2.6</td>
<td>Proposition</td>
<td>Mediangale exogeneity</td>
<td>6</td>
</tr>
<tr>
<td>3.2</td>
<td>Proposition</td>
<td>Sign distribution</td>
<td>9</td>
</tr>
<tr>
<td>3.3</td>
<td>Proposition</td>
<td>Randomized sign distribution</td>
<td>10</td>
</tr>
<tr>
<td>6.1</td>
<td>Assumption</td>
<td>Mixing</td>
<td>20</td>
</tr>
<tr>
<td>6.2</td>
<td>Assumption</td>
<td>Moment condition</td>
<td>20</td>
</tr>
<tr>
<td>6.3</td>
<td>Assumption</td>
<td>Boundedness</td>
<td>20</td>
</tr>
<tr>
<td>6.4</td>
<td>Assumption</td>
<td>Non-singularity</td>
<td>20</td>
</tr>
<tr>
<td>6.5</td>
<td>Assumption</td>
<td>consistent estimator of $V_n$</td>
<td>20</td>
</tr>
<tr>
<td>6.6</td>
<td>Theorem</td>
<td>Asymptotic distribution of statistic SHAC</td>
<td>20</td>
</tr>
<tr>
<td>6.8</td>
<td>Theorem</td>
<td>Generic Asymptotic Validity</td>
<td>21</td>
</tr>
<tr>
<td>6.9</td>
<td>Theorem</td>
<td>Monte Carlo test asymptotic validity</td>
<td>22</td>
</tr>
</tbody>
</table>
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, the parameters of the median regression are now better interpreted and 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.\footnote{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.} The most studied estimator is the Least Absolute Deviations (LAD) estimator. The actual 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...
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 linear 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)]. These methods, adapted to discrete statistics by a tie-breaking procedure [Dufour (2002)], permit to obtain exact simultaneous confidence region for \( \beta \). Then, conservative confidence intervals for each component of the parameter (or any real function 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 (1997b) extends the scheme of potential temporal dependence including ARMA disturbance processes. Horowitz (1998) uses a smoothed version of LAD. Thus, an important advance in this 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 (1997b) proposes moving block bootstrap. Finally, Hahn (1997) presents Bayesian bootstrap. The reader is referred to Buchinsky (1995, 2003), for a review and to Fitzenberger (1997b) 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 as-
ymptotic 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 remains 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 Dufouf (1981), Campbell and Dufouf (1995, 1997) and for $L_1$-estimation in Buchinsky (1995), De Angelis, Hall and Young (1993), Dielman and Pfaffenger (1988a, 1988b). 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 gather study of nonlinear functional forms and structural models with endogeneity ["censored quantile regressions", Powell (1984, 1986) and Fitzenberger (1997a), 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 $\beta$ 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 usual 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 $\beta$ by projection techniques [see Dufouf (2002)]. In a companion paper, we derive minimum distance estimators that optimize these statistics and study their properties and performance [see Coudin and Dufouf (2005)]. Once, 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 $W = \{W_t = (y_t, x'_t) : \Omega \rightarrow \mathbb{R}^{p+1}, t = 1, \ldots, n\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$. Let $\{\mathcal{F}_t, t \in \mathbb{N}\}$ be an adapted nondecreasing sequence of sub $\sigma$-fields of $\mathcal{F}$, i.e. $\mathcal{F}_t$ is a $\sigma$-field in $\Omega$ such that $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s < t$ and $\sigma(W_1, \ldots, W_t) \subseteq \mathcal{F}_t$, where $\sigma(W_1, \ldots, W_t)$ is the $\sigma$-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:

$$y_t = x'_t \beta + u_t, \quad t = 1, \ldots, n, \quad (2.1)$$

or, in vector notation,

$$y = X \beta + u, \quad (2.2)$$

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

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

$$E(u_t \mid \mathcal{F}_{t-1}) = 0, \quad \forall t \geq 1.\quad (2.3)$$

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 \( \{F_t, \ t = 1, 2, \ldots \} \) iff
\[
P[u_1 < 0] = P[u_1 > 0] = 0.5, \]
\[
P[u_t < 0|F_{t-1}] = P[u_t > 0|F_{t-1}] = 0.5, \text{ for } t > 1. \]

**Definition 2.3**  STRICT CONDITIONAL MEDIANGALE. Let \( \{u_t, F_t, \ t = 1, 2, \ldots \} \) be an adapted sequence and \( F_t = \sigma(u_1, \ldots, u_t, X) \). Then \( \{u_t, \ t = 1, 2, \ldots \} \) is a strict mediangale conditional on \( X \) with respect to \( \{F_t, \ t = 1, 2, \ldots \} \) 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 \( \{u_t, F_t, \ t = 1, 2, \ldots \} \) be an adapted sequence and \( F_t = \sigma(u_1, \ldots, u_t, X) \). Then \( \{u_t, \ t = 1, 2, \ldots \} \) is a weak mediangale conditional on \( X \) with respect to \( \{F_t, \ t = 1, 2, \ldots \} \) iff
\[
P[u_1 > 0|X] = P[u_1 < 0|X],
\]
\[
P[u_t > 0|u_1, \ldots, u_{t-1}, X] = P[u_t < 0|u_1, \ldots, u_{t-1}, X], \text{ for } t = 2, \ldots, n. \]

Stating that \( \{u_t, \ t = 1, 2, \ldots \} \) is a weak mediangale with respect to \( \{F_t, \ t = 1, 2, \ldots \} \) 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-\( \sigma \) algebras \( \{F_t, \ t = 1, 2, \ldots \} \). However, the weak mediangale concept is slightly more demanding than usual statement of martingale difference of signs. Indeed, the sequence sub-\( \sigma \) algebras of reference is usually taken to be \( \{F_t = \sigma(W_1, \ldots, W_t), \ t = 1, 2, \ldots \} \) and \( \{s(u_t) \odot x_t, F_t \} \) is assumed to be martingales with respect to \( \sigma(W_1, \ldots, W_t), \ t = 1, 2, \ldots \). Here, the sequence sub-\( \sigma \) algebras of reference is \( \{F_t = \sigma(W_1, \ldots, W_t, X), \ t = 1, 2, \ldots \} \). Conditional mediangale requires conditioning on the whole process \( X \). We shall see later that asymptotic inference may be available under weaker assumptions, as a mediangale 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:

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

Our last remark concerns exogeneity. As long as the \( x_t \)'s are strictly exogenous explanatory vari-
ables, conditional mediangale concept is equivalent to usual martingale difference for signs with
respect to \( F_t = \sigma(W_1, \ldots, W_t), \ t = 1, 2, \ldots \).

**Proposition 2.6** **Mediagale Exogeneity.** Suppose \( \{x_t: t = 1, \ldots, n\} \) is a strictly exoge-
nous process for \( \beta \) 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 (1997b), 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. 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.

---

In Fitzenberger (1997b), LAD and quantile estimators are shown to be consistent and asymptotically normal if
amongst other, \( E[x_t \delta_0(u_t)] = 0, \ \forall t = 1, \ldots, n, \) densities exist and second order moments for \( (u_t, x_t) \) are finite.
2.2. Special cases

The above framework obviously covers a large spectrum of heteroskedasticity 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

$$u_1 = \sigma_1(x_1, \ldots, x_{t-1}) \varepsilon_1,$$

$$u_t = \sigma_t(x_1, \ldots, x_{t-1}, u_1, \ldots, u_{t-1}) \varepsilon_t, \ t = 2, \ldots, n$$

where \( \varepsilon_1, \ldots, \varepsilon_n \) are independent with median 0.

$$P[u_t > 0] = P[u_t < 0] = \frac{1}{2}, \ t = 2, \ldots, n,$$

\( \sigma_1(x_1, \ldots, x_{t-1}) \) and \( \sigma_t(x_1, \ldots, x_n, u_1, \ldots, u_{t-1}) \), \( t = 2, \ldots, n \),

are non-zero with probability one.

In time series context, this includes:

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

$$u_t = \exp(w_t/2)r_y\varepsilon_t,$$

$$w_t = a_1 w_{t-1} + \cdots + a_1 w_{t-p} + r_w v_t,$$

\( v_1, \ldots, v_n \) are i.i.d. random variables.

The first two examples are especially relevant for cross-sectional data where procedures and estimators are expected to be robust to heterogeneity. Other examples present robustness properties to endogenous disturbances variance (or volatility) specification. 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]. This property is more general and does not specify explicitly the functional
form of the variance as in an ARCH specification. In microeconometrics, such assumption may be relevant for panel data or stratified cross-sections.

3. **Exact finite-sample sign-based inference**

3.1. **Motivation**

In econometrics, tests are often based on $t$ or $\chi^2$ 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 $\beta$ 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 Pfafffenerberger (1988a, 1988b). 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 hypothesis:

$$H_0 : \ X_1, \ldots, X_n \text{ are independent observations}\quad$$

where each one with a distribution symmetric about zero. \hfill (3.1)

Here, $H_0$ allows for arbitrary heteroskedasticity. Let

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

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

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

$$P[\text{Rejecting } H_0 \mid |X_1|, \ldots, |X_n|] \leq \alpha \text{ under } H_0. \quad (3.3)$$

This theorem directly implies that all procedures typically designated as "robust to heteroskedasticity" or "HAC" [see White (1980), Newey and West (1987), Andrews (1991), etc..], which do not satisfy condition 3.3, have size one. 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.

### 3.2. 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

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

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

\[
s(y - X\beta) = [s(y_1 - x'_1\beta), \ldots, (y_n - x'_n\beta)]'
\]

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

\[
T = T[s(y - X\beta), X]
\]

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.

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 \(X\),

\[
T[s(y - \beta_0X), X] \sim T(S_n, X)
\]

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 $\alpha$ rejects the null hypothesis when

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

(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:

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

(3.9)

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

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

(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:

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

(3.11)

or, equivalently,

$$\tilde{s}(x, V) = s(x), \quad \text{if} \quad s(x) = +1 \text{ or } -1$$
$$s(V - 0.5), \quad \text{if} \quad s(x) = 0.$$  

(3.12)

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

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

(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

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

(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 \( \beta \) 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:

\[
D_S(\beta_0, \Omega_n) = s(y - X\beta_0)'X\Omega_n[s(y - X\beta_0), X]X's(y - X\beta_0)
\]  
(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

\[
SB(\beta_0) = s(y - X\beta_0)'XX's(y - X\beta_0) = ||X's(y - X\beta_0)||^2
\]  
(4.16)

and

\[
SF(\beta_0) = s(y - X\beta_0)'P(X)s(y - X\beta_0) = ||X's(y - X\beta_0)||^2_M
\]  
(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 \( \beta \in \mathbb{R}^p \) 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 mediagale property and for which the conditional density at zero is the same \( f_i(0|X) = f(0), \forall t = 1, \ldots, n \). \( SF(\beta_0) \) 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 \( X \):

\[
s(y - X\beta_0) = X\gamma + v.
\]  
(4.18)

\( SF(\beta_0) \) may also be interpreted as the Mahalanobis norm of \( s(y - X\beta_0) \).

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 mediagale 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

\[
SB(\beta_0) = s(y - X\beta_0)'\tilde{X}\tilde{X}'s(y - X\beta_0)
\]  
(4.19)
where

\[
\hat{X} = \left\{ \begin{array}{cccc}
  f_1(0|X) & 0 & \ldots & 0 \\
  0 & f_1(0|X) & \ldots & f_n(0|X) \\
  \ldots & \ldots & \ldots & \ldots \\
  0 & 0 & \ldots & f_n(0|X)
\end{array} \right\} X.
\]

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. \( \hat{X} \) 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

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

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

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

\( J_n[s(y - X\beta_0), X] \) 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

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

12
where

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

(4.23)

and the \( V_t \) process has spectral density matrix

\[
f(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{+\infty} \Gamma(j)e^{-ij\lambda}, \text{ with } \Gamma(j) = E[V_tV_{t-j}].
\]  

(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

\[
\hat{J}_n = \frac{n}{n-p} \sum_{j=-n+1}^{n-1} k\left(\frac{j}{B_n}\right)\hat{\Gamma}_n(j),
\]  

(4.25)

where

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

(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:

\[
SB(\beta_0) > c_{SB}(X, \alpha),
\]

\[
SF(\beta_0) > c_{SF}(X, \alpha),
\]

\[
DS(\beta_0, \hat{J}_n^{-1}) > c_{SHAC}(X, \alpha),
\]

where \( c_{SB}(X, \alpha), c_{SF}(X, \alpha) \) and \( c_{SHAC}(X, \alpha) \) are critical values which depend on the level \( \alpha \) 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, see 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 replications of $T$. The empirical $p$-value is given by

$$\hat{p}_N(x) = \frac{NG_N(x) + 1}{N + 1}$$

(4.27)

where

$$G_N(x) = \frac{1}{N} \sum_{i=1}^{N} 1_{[0, \infty)}(T_i - x), \quad 1_A(x) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \not\in A. \end{cases}$$

Then we have

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

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, this 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:

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

The adapted empirical p-value is given by

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

where

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

Then

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

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 $\alpha$, 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 replications, such that $\alpha(N + 1)$ is an integer, where $\alpha$ is the desired level. Then, we generate $N$ replications $SF_j = S_j'X(X'X)^{-1}X'S_j$ where $S_j$ is a realization of a $n$-vector of independent Bernoulli random variables, and we compute $\hat{p}_N(SF(\beta_0))$. Finally, $H_0(\beta_0)$ is rejected at level $1 - \alpha$ if $\hat{p}_N(SF(\beta_0)) < \alpha$.

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 $\beta$ 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 $\beta = \beta_0$ [see Schweder and Hjort (2002)]. As we will see later, that idea is clearly related to the notion of confidence distribution when $\beta$ is of dimension one.

The confidence set with level $1 - \alpha$, $C_\beta(\alpha)$ is the set of all $\beta_0$ with $p$-value higher than $1 - \alpha$. By construction, $C_\beta(\alpha)$ has level $1 - \alpha$. From this simultaneous confidence set for $\beta$, 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 $\beta$. For examples of use in different settings and for further discussion, see 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 $\beta$ is available. Its orthogonal projections on each axis of $\mathbb{R}^p$ will give confidence intervals for each component $\beta_k$ of size larger than $1 - \alpha$. To be more precise, suppose $C_\beta(\alpha)$ is a confidence set with level $1 - \alpha$ for $\beta$:

$$P[\beta \in C_\beta(\alpha)] \geq 1 - \alpha. \quad (5.1)$$

Since

$$\beta \in C_\beta(\alpha) \implies g(\beta) \in g(C_\beta(\alpha)), \quad (5.2)$$
we have:
\[ P[\beta \in C_\beta(\alpha) | \geq 1 - \alpha \implies P[g(\beta) \in g(C_\beta(\alpha))] \geq 1 - \alpha. \]

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

Hence, to obtain valid conservative confidence intervals for the component \( \beta_k \) of the \( \beta \) parameter in
the model (2.1) under mediagale Assumption 2.5, it is sufficient to solve both following numerical optimization problems, (stated here for the statistic \( SF \)),
\[
\min_{\beta \in \mathbb{R}^p} \beta_k \\
\text{s.t. } \tilde{p}_N(SF(\beta)) \geq 1 - \alpha
\]
\[
\text{and}
\max_{\beta \in \mathbb{R}^p} \beta_k \\
\text{s.t. } \tilde{p}_N(SF(\beta)) \geq 1 - \alpha
\]

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, using a global search optimization algorithm, like \textbf{simulated annealing algorithm}, see 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 \( \beta_0 \) of \( \beta \). This \( p \)-value can be seen as a sort of degree of confidence one may have in the value \( \beta_0 \). By aggregating values of \( \beta \) with given \( p \)-values, we can construct simultaneous confidence regions for any significance level of coverage.

In the special case when \( \beta \) has dimension one, the function that associates the simulated \( p \)-value to each value \( \beta_0 \) of \( \beta \) is related to a \textbf{confidence distribution} of \( \beta \) given the realizations \((y, X)\) [see 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 \( \beta \) is defined as a distribution with cumulative \( CD(\beta) \) and quantile
function $CD^{-1}(\beta)$, such that

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

(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 $\alpha$.\(^3\) 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

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

(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 $\beta$ and follows a uniform distribution, $CD(\beta)$ constitutes a pivot conditional on $X$. Reciprocally, whenever a pivot is available for example any statistics $T(\beta)$ that increases with $\beta$ and that has a cumulative distribution function $F$ independent of $\beta$ and of any nuisance parameter, $F[T(\beta)]$ is uniformly distributed and thus is 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 $\beta$. They can, in a sense, be compared to Bayesian posterior probabilities for a frequentist setup. To any value of $\beta$, 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 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 $\beta$ 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 $n = 50$,

$$y_t = \beta_0 + \beta_1 x_t + u_t, \quad t = 1, \ldots, n,$$

(5.6)

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

\(17\)
\( u_t \overset{i.i.d.}{\sim} \begin{cases} N[0, 1] & \text{with probability 0.95} \\ N[0, 100^2] & \text{with probability 0.05} \end{cases} \)

\( \beta_0 = \beta_1 = 0. \)

We conduct an exact inference procedure. As \( \beta \) is a 2-vector, we can have a graphic illustration. To each value of the vector \( \beta \) is associated a \( p \)-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 \( \beta \) where \( p \)-values are bigger than \( 1 - \alpha \). Same graphics for other estimators are provided in appendix.

Confidence distribution of SF-based inference
Confidence sets for Mahalanobis sign-test approach

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 [see Coudin and Dufour (2005) for the analyze of sign-based estimation].

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>White</th>
<th>SF</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>-1.876</td>
<td>-1.876</td>
<td>-0.28</td>
</tr>
<tr>
<td>(s.d.)</td>
<td>(1.343)</td>
<td>(1.293)</td>
<td>-</td>
</tr>
<tr>
<td>90%CI</td>
<td>[-4.126, 0.375]</td>
<td>[-4.043, 0.292]</td>
<td>[-0.52, 0.17]</td>
</tr>
<tr>
<td>95%CI</td>
<td>[-4.574, 0.822]</td>
<td>[-4.474, 0.722]</td>
<td>[-0.54, 0.23]</td>
</tr>
<tr>
<td>98%CI</td>
<td>[-5.103, 1.351]</td>
<td>[-4.983, 1.232]</td>
<td>[-0.64, 0.26]</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.358</td>
<td>0.358</td>
<td>0.14</td>
</tr>
<tr>
<td>(s.d.)</td>
<td>(1.424)</td>
<td>(0.845)</td>
<td>-</td>
</tr>
<tr>
<td>90%CI</td>
<td>[-2.030, 2.745]</td>
<td>[-1.058, 1.774]</td>
<td>[-0.36, 0.46]</td>
</tr>
<tr>
<td>95%CI</td>
<td>[-2.504, 3.219]</td>
<td>[-1.340, 2.055]</td>
<td>[-0.42, 0.59]</td>
</tr>
<tr>
<td>98%CI</td>
<td>[-3.065, 3.780]</td>
<td>[-1.673, 2.388]</td>
<td>[-0.57, 0.64]</td>
</tr>
</tbody>
</table>

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 for tests. We point out that mediangale Assumption 2.5 can be seen as too restrictive and excludes some processes whereas asymptotic inference still
can be conducted on these processes. 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 $D_S(\beta_0, \hat{J}_n^{-1})$ 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. $\{(x'_t, s(u_t))\}$ is $\alpha$-mixing of size $-2r/(r-2)$ with $r > 2$.\(^4\)

**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. $V_n = \text{var}[\frac{1}{\sqrt{n}} \sum_{t=1}^{n} s(u_t)x_t]$ is uniformly positive definite.

**Assumption 6.5** Consistent estimator of $V_n$. $\Omega_n(\beta_0)$ is symmetric positive definite uniformly over $n$ and $\Omega_n - V_n \rightarrow_p 0$.

**Theorem 6.6** Asymptotic distribution of statistic SHAC. In model (2.1), with conditions 6.1-6.5, we have, under $H_0$,

$$D_S(\beta_0, \hat{J}_n^{-1}) \rightarrow \chi^2(p).$$

**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 $H_0$,

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

When the mediangale condition holds, $J_n$ reduces to $E(X'X/n)$. Hence, $(X'X/n)^{-1}$ is a consistent estimator of $J_n^{-1}$. In the same context, $SB(\beta_0)$ converges in distribution to a non-central $\chi^2$ distribution.

\(^4\)Mixing concept is defined in annex
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 \( \{ M_\beta, \beta \in \Theta \} \), where the parameter \( \beta \) 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 \( S^n_0(\beta_0) \) be the observed statistic and \( c_n \) the rate of convergence. We suppose that, under \( H_0(\beta_0) \), \( c_nS^n_0 \) 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

\[
\tilde{G}_n(x|X_n(\omega)) \rightarrow G(x), \text{ with probability one,}
\]

where \( G \) does not depend on the realization \( X(\omega) \). The distribution of \( c_nS^n_0(\beta_0) \) 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 \( H_0(\beta_0) \),

\[
P[c_nS^n_0 \geq x|X_n] = G_n(x|X_n) = 1 - F_n(x|X_n) \rightarrow G(x) \text{ a.e.,}
\]

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

\[
\tilde{G}_n(x|X_n(\omega)) \rightarrow G(x) \text{ a.e. with probability one.}
\]

Then

\[
\lim_{n \rightarrow \infty} P[\tilde{G}_n(c_nS^n_0, X_n(\omega)) \leq \alpha] \leq \alpha. \tag{6.1}
\]

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 \( U[0, 1] \) distribution, \( S^n_0 \) the observed statistics and \( S^n(N) = (S^n_1, \ldots, S^n_N) \), \( N \) independent replications drawn from \( F_n(x) \). Then, the randomized empirical survival function under the null hypothesis is

\[
\tilde{G}^n_{N,n}[x, n, U_0, S^n_0, S^n(N), U(N)] = 1 - \frac{1}{N} \sum_{j=1}^{N} u(x - c_nS^n_j) + \frac{1}{N} \sum_{j=1}^{N} \delta(c_nS^n_j - x)u(U_j - U_0)
\]

with, \( u(x) = 1_{[0, \infty)}(x), \delta(x) = 1_{\{0\}} \). Note that \( \tilde{G}^n_{N,n}[x, S^n_0, S^n(N), n] \) is in a sense an approximation of \( \tilde{G}_n(x) \), and thus depends on \( N \) the number of replications and \( n \) the number of observations.
The randomized empirical $p$-value function is defined as
\[ \tilde{p}_N^r(x) = \frac{N \tilde{G}_{N,n}(x) + 1}{N + 1}. \] (6.3)

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

**Theorem 6.9  Monte Carlo Test Asymptotic Validity.** Let $S_n(0) = S_0^n$ 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 $H_0(\beta_0)$,
\[ P[c_n S_0^n \geq x | X_n] = G_n(x | X_n) = 1 - F_n(x | X_n) \rightarrow G(x) \text{ a.e.}, \]
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
\[ P[c_n S_n \geq x | X_n] = \tilde{G}_n(x | X_n) = 1 - \tilde{F}_n(x | X_n) \rightarrow G(x) \text{ a.e.}, \]
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 $\alpha$ is asymptotically valid, i.e.
\[ \lim_{n \to \infty} P[\tilde{p}_N^r(\beta_0) \leq \alpha] = \alpha. \] (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 function may be different from a Gaussian distribution. 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 $SF(\beta)$ be the test statistic as defined in (4.21).

- Observe $SF(\beta_0)$. Draw $N$ replications 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” replications of $SF(\beta_0)$ under the null. We call them "pseudo" replications because they are drawn as if observations were independent, which may be not the case for the observed $SF(\beta_0)$. 

22
• Draw \( N + 1 \) independent replications \((W_0, \ldots, W_N)\) from a \( U_{[0,1]} \) distribution and form the couple \((SF^j, W_j)\), where \( SF^0 \) stands for the observed \( SF(\beta_0) \).

• Compute \( \hat{p}_N(\beta_0) \) using (6.3).

• From Theorem 6.9, the confidence region \( \{ \beta \in \mathbb{R}^p | \hat{p}_N(\beta) \geq \alpha \} \) is asymptotically conservative with level at least \( 1 - \alpha \). We reject \( H_0 \) if \( \hat{p}_N(\beta_0) \leq \alpha \).

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 (1997b) 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 of estimators that may follow from the conditional density at 0 of the process \( \{ u \} \).

7. Simulations and comparisons with other methods

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

\[
y_t = x_t' \beta_0 + u_t,
\]

where \( x_t = (1, x_{2,t}, x_{3,t})' \) and \( \beta_0 \) are \( 3 \times 1 \) 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 11 general DGP. For the first 8 ones, \( u_t \) may depend on the explanatory variables and its past in a nonlinear heteroskedastic or dependent way,

\[
u_t = h(x_t, u_{t-1}, \ldots, u_1) \epsilon_t,
\]

provided that \( u \) constitutes a strict conditional mediangale given \( X \) (see Assumption 2.5). In these cases, the sign-based inference leads to exact levels. For the last 4 ones, we study the behavior of the sign-based inference (involving a HAC correction) when it is only asymptotically valid. In those
cases, $x_t$ and $u_t$ are such that $E(x_t u_t) = 0$ and $E(x_t \text{sign}(u_t)) = 0$.

Throughout these DGP, we illustrate different classical problems that may be encountered in practice. For cases 1-3, 9-11, we employ the following stochastic processes (see Fitzenberger (1997a)).

**AR(1)-HOM:**

$$x_{j,t} = \rho_x x_{j,t-1} + \nu_{j,t},$$

$$u_t = \rho_u u_t + \nu_{t}^u,$$

**AR(1)-HET:**

$$x_{j,t} = \rho_x x_{j,t-1} + \nu_{j,t},$$

$$\tilde{u}_t = \rho_u \tilde{u}_t + \nu_{t}^u,$$

and $$u_t = \min\{3, \max[0.21, |x_{2,t}|]\} \times \tilde{u}_t$$

where the innovations $\nu_{t}^u$ and $\nu_{j,t}$ are centered normal with variance 1.

**CASE 1:** $\rho_x = \rho_u = 0$, **HOM**.

**CASE 2:** $\rho_x = \rho_u = 0$, **HET**.

**CASE 3:** $\rho_x = 0.5, \rho_u = 0$, **HET**.

**CASE 9:** $\rho_x = 0, \rho_u = 0.5$, **HOM**.

**CASE 10:** $\rho_x = 0.5, \rho_u = 0.5$, **HET**.

**CASE 11:** $\rho_x = 0, \rho_u = 0.9$, **HOM**.

For cases 4-8, we study other kinds of heteroskedasticity and nonlinear dependence (GARCH, Stochastic Volatility ...), debalanced scheme in the explanatory variables and heavy-tailed distributions for the error term.

**CASE 4:**

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

$$u_t = \exp(2 \times \epsilon_t),$$

where $\epsilon_t \sim N(0,1)$ and are i.i.d..

**CASE 5:** Stochastic Volatility:

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

$$u_t = \exp(u_{t-1}/2) \epsilon_t,$$

$$w_t = 0.5 w_{t-1} + v_t,$$

where $\epsilon_t \sim N(0,1), v_t \sim \chi^2(3)$ and are i.i.d.
CASE 6: GARCH(1,1):
\[ y_t = x_t^\prime \beta_0 + u_t, \]
\[ u_t = \sigma_t \epsilon_t, \]
\[ \sigma_t^2 = 0.8u_{t-1}^2 + 0.8\sigma_{t-1}^2, \]
where \( \epsilon_t \sim \mathcal{N}(0, 1) \) and are i.i.d..

CASE 7: DEB1:
\[ y_t = x_t^\prime \beta_0 + u_t, \]
where \( x_{2,t} \sim \mathcal{B}(1, 0.3), \]
\( x_{3,t} \sim \mathcal{N}(0, 0.1^2), \]
\( u_t \sim \mathcal{N}(0, 1) \) and are i.i.d..

CASE 8: CAUCHY:
\[ y_t = x_t^\prime \beta_0 + u_t, \]
where \( x_t \sim \mathcal{N}(0, I_2) \),
\( u_t \sim \mathcal{C} \) and are i.i.d..

Finally, cases 1 and 2 present iid normal observations without and with conditional heteroskedasticity. Cases 3 involves nonlinear dependence in the error term. Cases 4, 5 and 6 are other cases of heteroskedasticity and nonlinear dependence in the error term, where Assumption 2.5 can hold. DGP 7 and 8 present very debalanced schemes in the design matrix (a case when the LAD estimator is known to perform badly). Case 9 is an example of long tailed errors. We also investigated the behavior of sign-based inference when a linear dependence is present in the error term. Cases 9 to 11 illustrate different levels of autocorrelation with and without heteroskedasticity.

7.1. Inference
We first study inference. 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 \( D_S(\hat{\beta}, \hat{J}_n^{-1}) \) 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 for 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 \(\theta\)th order statistics [see Buchinsky (2003)].

• BARTLAD: Bartlett kernel estimator for LAD asymptotic covariance matrix [see Powell (1984), Fitzenberger (1997b), 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 (1997b)].

• 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 3, 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. 4999. All bootstrapped samples are of size \(T = 50\). We perform \(S = 1000\) simulations to estimate the level of these tests. In cases 1-8, one can exactly control the level of the test with a sign-based method. Remark that, for cases involving a strong heteroskedasticity (4-6), asymptotic tests greatly under reject the null hypothesis. Thus, we may expect these tests to have very poor power for various alternatives. In those cases, sign-based methods keep their good properties. When linear dependence is present in the data, sign-based methods loose their exactness. However, with a HAC correction 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 \(\beta_1\) and \(\beta_3\) at the tested value, i.e \(\beta_1 = 1\) and \(\beta_3 = 3\), and letting vary \(\beta_2\). Simulated power is given by a graph with \(\beta_2\) in abscissa. The power functions presented here are locally adjusted for the level. 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.
Table 3: Level of Conditional Tests.

<table>
<thead>
<tr>
<th>Case</th>
<th>$\rho_e = \rho_x = 0$, HOM</th>
<th>$\rho_e = \rho_x = 0$, HET</th>
<th>$\rho_e = 0$, $\rho_x = .5$, HET</th>
<th>$u_t = \exp(2t)\epsilon_t$</th>
<th>Stochastic Volatility, GARCH(1,1),</th>
<th>Stochastic Volatility, $x_{2t} \sim B(1, 1/3)$, $x_{3t} \sim N(0,.01^2)$</th>
<th>$(x_{2t}, x_{3t}) \sim N(0, I_2)$, $u_t \sim C$</th>
<th>$\rho_e = .5$, $\rho_x = 0$, HOM</th>
<th>$\rho_e = \rho_x = .5$, HET</th>
<th>$\rho_e = .9$, $\rho_x = 0$, HOM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\text{sign}$</td>
<td>$\text{sign}$</td>
<td>$\text{LAD}$</td>
<td>$\text{LAD}$</td>
<td>$\text{LAD}$</td>
<td>$\text{OLS}$</td>
<td>$\text{OLS}$</td>
<td>$\text{OLS}$</td>
<td>$\text{OLS}$</td>
<td>$\text{OLS}$</td>
</tr>
<tr>
<td></td>
<td>SF</td>
<td>HAC</td>
<td>OS</td>
<td>DMB</td>
<td>MBB5</td>
<td>BART</td>
<td>IID</td>
<td>WH</td>
<td>BART</td>
<td></td>
</tr>
<tr>
<td>Case 1:</td>
<td>.957</td>
<td>.957</td>
<td>.914</td>
<td>.960</td>
<td>.914</td>
<td>.949</td>
<td>.946</td>
<td>.899</td>
<td>.888</td>
<td></td>
</tr>
<tr>
<td>Case 2:</td>
<td>.954</td>
<td>.952</td>
<td>.739</td>
<td>.973</td>
<td>.933</td>
<td>.944</td>
<td>.836</td>
<td>.890</td>
<td>.860</td>
<td></td>
</tr>
<tr>
<td>Case 3:</td>
<td>.953</td>
<td>.947</td>
<td>.684</td>
<td>.962</td>
<td>.935</td>
<td>.933</td>
<td>.775</td>
<td>.887</td>
<td>.866</td>
<td></td>
</tr>
<tr>
<td>Case 4:</td>
<td>.948</td>
<td>.943</td>
<td>.886</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>.962</td>
<td>.981</td>
<td>.981</td>
<td></td>
</tr>
<tr>
<td>Case 5:</td>
<td>.958</td>
<td>.962</td>
<td>.955</td>
<td>.995</td>
<td>.988</td>
<td>.997</td>
<td>.966</td>
<td>.988</td>
<td>.988</td>
<td></td>
</tr>
<tr>
<td>Case 6:</td>
<td>.959</td>
<td>.959</td>
<td>.981</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>.895</td>
<td>.992</td>
<td>.980</td>
<td></td>
</tr>
<tr>
<td>Case 7:</td>
<td>.955</td>
<td>.943</td>
<td>.931</td>
<td>.969</td>
<td>.906</td>
<td>.939</td>
<td>.893</td>
<td>.935</td>
<td>.854</td>
<td></td>
</tr>
<tr>
<td>Case 8:</td>
<td>.945</td>
<td>.950</td>
<td>.930</td>
<td>.980</td>
<td>.961</td>
<td>.984</td>
<td>.967</td>
<td>.978</td>
<td>.978</td>
<td></td>
</tr>
<tr>
<td>Case 9:</td>
<td>.857</td>
<td>.975</td>
<td>.840</td>
<td>.888</td>
<td>.893</td>
<td>.899</td>
<td>.821</td>
<td>.785</td>
<td>.700</td>
<td></td>
</tr>
<tr>
<td>Case 10:</td>
<td>.787</td>
<td>.967</td>
<td>.425</td>
<td>.869</td>
<td>.891</td>
<td>.829</td>
<td>.516</td>
<td>.616</td>
<td>.680</td>
<td></td>
</tr>
<tr>
<td>Case 11:</td>
<td>.911</td>
<td>.988</td>
<td>.860</td>
<td>.918</td>
<td>.932</td>
<td>.894</td>
<td>.860</td>
<td>.803</td>
<td>.736</td>
<td></td>
</tr>
</tbody>
</table>

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. In very heteroskedastic cases (5, 6), sign-based inference greatly dominates other methods: levels of coverage are exactly controlled and the power function is largely higher than for other methods, even with locally adjusted levels. When autocorrelation is present in the data, the sign-based inference looses the exactness of the level. However, when a HAC correction is included in the sign statistic, the method seems to lead to comparable performance than usual methods such as kernels, moving block bootstrap. Only for very high autocorrelation (close to unit root process), the sign-based inference is not adapted.
7.2. Confidence intervals

As the sign-based confidence regions are by construction of coverage level higher that 1 − α whenever inference is exact. We consider as indicator of performance their spread and more precisely the spread of projection-based confidence intervals. - TO BE CONTINUED-
8. Conclusion

In this paper, we have proposed valid sign-based inference procedures for the \( \beta \) 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.
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 \([s(u_1), s(u_2), \ldots, s(u_n)]' = (s_1, s_2, \ldots, s_n)'.\) 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
\[
P[u_t > 0 | u_{t-1}, \ldots, u_1, X] = 1/2 = P[u_t > 0 | u_{t-1}, \ldots, s_1, X], \forall t \geq 2.
\]
Further, the joint density of \((s_1, s_2, \ldots, s_n)'\) can be written:
\[
l(s_1, s_2, \ldots, s_n | X) = \prod_{t=1}^n l(s_t | s_{t-1}, \ldots, s_1, X)
\]
\[
= \prod_{t=1}^n P[u_t > 0 | u_{t-1}, \ldots, u_1, X]^{(1-s_t)/2} \{1 - P[u_t > 0 | u_{t-1}, \ldots, u_1, X]\}^{(1+s_t)/2}.
\]
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:
\[
\tilde{s}(u_t, V_t) = s(u_t) + [1 - s(u_t)] s(V_t - .5).
\]
Hence
\[
P[\tilde{s}(u_t, V_t) = 1 | u_{t-1}, \ldots, u_1, X] = P[s(u_t) + [1 - s(u_t)] s(V_t - .5) | u_{t-1}, \ldots, u_1, X].
\]
As \((V_1, \ldots, V_n)\) is independent of \((u_1, \ldots, u_n)\) and \(U(0, 1)\) 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 mediagale 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

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

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 mediagale 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 \(\alpha\), 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

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

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

\[
P[S(y) = s|\beta, X] = \prod_{i=1}^{n} [P(y_i > 0|\beta, X)]^{(1+s_i)/2}[P(y_i < 0|\beta, X)]^{(1-s_i)/2} \tag{A.6}
\]

\[
= \frac{1}{2^n} \prod_{i=1}^{n} [1 + 2f_i(0|X)x_is_i\beta + o(\beta)] \tag{A.7}
\]

\[
= \frac{1}{2^n} [1 + 2 \sum_{i=1}^{n} f_i(0|X)x_is_i\beta + o(\beta)]. \tag{A.8}
\]
And by identification,

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

(A.9)

Therefore, the required test has the form

\[
W = \{ s = (s_1, \ldots, s_n) \left| \sum_{i=1}^{n} f_i(0|X) x_i s_i > c_\alpha \right. \},
\]  

(A.10)

or equivalently,

\[
W = \{ s|s'(y)X'X(y) > c'_\alpha \},
\]  

(A.11)

where \( c_\alpha \) and \( c'_\alpha \) 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,

\[
W = \{ s|s'(y)X'X(y) > c'_\alpha \}.
\]  

(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,

\[
\frac{dP[W|\beta]}{d\beta} \bigg|_{\beta=0} = 0,
\]  

(A.13)

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

\[
\beta \cdot \frac{1}{2} \left( \frac{d^2 P[W|\beta]}{d\beta^2} \right) \beta = \frac{1}{2n} \sum_{1 \leq i \neq j \leq n} \sum_{k=1}^{p} f_i(0|X) s_i \beta' x_i [f_j(0|X) s_j \beta' x_i].
\]  

(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 \( \frac{d^2 P[W|\beta]}{d\beta^2} \) at \( \beta = 0 \) [see Boldin et al. (1997)]. Taking the trace in expression (A.14), we find (after some computations) it is proportional to

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

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

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

\[
W = \{ s : s'(y) \bar{X} \bar{Y}s(y) > c' \alpha \}.
\] (A.17)

The same reasoning with another definition of curvature leads to

\[
W = \{ s : s'(y) \bar{X}(\bar{Y})^{-1} \bar{Y}s(y) > c' \alpha \}.
\] (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 \( Z_{nt} = \lambda' V_n^{-1/2} x_t s(u_t) \), for some \( \lambda \in \mathbb{R}^p \) such that \( \lambda' \lambda = 1 \). The mixing property of \( (x_t, \epsilon_t) \) (condition 6.1) is reported on \( Z_{nt} \) [Theorem 3.49, in White (2001)]. Hence, \( \lambda' V_n^{-1/2} s(u_t) \otimes x_t \) is \( \alpha \)-mixing of size \( -r/(r-2) \), \( r > 2 \).

2. Condition 6.2 implies

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

Condition 6.3 implies

\[
E[\lambda' V_n^{-1/2} x_t s(u_t)]^\top < \Delta < \infty, \quad \forall t = 1, \ldots, n, \quad \forall n \in \mathbb{N}.
\] (B.20)

Remark that

\[
\text{var}\left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Z_{nt} \right) = \text{var}\left[ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \lambda' V_n^{-1/2} s(u_t) \otimes x_t \right] = \lambda' V_n^{-1/2} V_n V_n^{-1/2} \lambda = 1.
\] (B.21)

3. The mixing property of \( Z_{nt} \) 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

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \lambda' V_n^{-1/2} s(u_t) \otimes x_t \to \mathcal{N}(0, 1).
\] (B.22)
4. Since \( \lambda \) can be arbitrary chosen provided that \( \lambda' \lambda = 1 \), Cramér-Wold device implies
\[
V_n^{-1/2} n^{-1/2} \sum_{t=1}^{n} s(u_t) \otimes x_t \rightarrow \mathcal{N}(0, I_p).
\]

5. Finally, condition 6.5 states that \( \Omega_n \) is a consistent estimate of \( V_n^{-1} \). Hence,
\[
n^{-1/2} \Omega_n^{1/2} \sum_{t=1}^{n} s(u_t) \otimes x_t \rightarrow \mathcal{N}(0, I_p).
\]

Under \( H_0 \), \( s_t = s(y_t - x_t' \beta_0) \). Finally,
\[
n^{-1} s'(y - X\beta_0)X\Omega_nX's(y - X\beta_0) \rightarrow \chi_2(p).
\]

B.2. Proof of Corollary 6.7
Let \( F_t = \sigma(y_0, \ldots, y_t, x_0', \ldots, x_t') \). When the mediangale Assumption 2.5 holds, \( \{s(u_t) \otimes x_t, F_t, t = 1, \ldots, n\} \) result from a martingale difference with respect to \( F_t \). Hence,
\[
V_n = \text{Var}[\frac{1}{\sqrt{n}}s \otimes X] = \frac{1}{n} \sum_{t=1}^{n} E(x_t s_t s_t') = \frac{1}{n} \sum_{t=1}^{n} E(x_t x_t') = \frac{1}{n} E(X'X),
\]
and \( X'X/n \) is a consistent estimate of \( E(X'X/n) \). Theorem 6.6 gives \( SF(\beta_0) \rightarrow \chi_2(p) \). □

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,
\[
F_n(x) \rightarrow_{n \rightarrow \infty} F(x), \ \forall x \in \mathbb{R}.
\]

Then, \( (F_n) \) converges uniformly to \( F \) in \( \mathbb{R} \), i.e.
\[
\sup_{-\infty < x < +\infty} |F_n(x) - F(x)| \rightarrow_{n \rightarrow \infty} 0.
\]

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 $k$:
\[ |F_{n_k}(x_k) - F(x_k)| \geq \epsilon > 0 \] (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 \( \xi \) 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 \( x_k < \xi \):
  \[
  \epsilon \leq F_{n_k}(x_k) - F(x_k) \leq F_{n_k}(\xi^-) - F(r_{m}^{a}) \\
  \leq F_{n_k}(\xi^-) - F_{n_k}(\xi) + F_{n_k}(r_{m}^{b}) - F(r_{m}^{b}) - F(r_{m}^{a}).
  \]

- **Case 2:** \( \{x_k\} \) is increasing and \( x_k < \xi \):
  \[
  \epsilon \leq F(x_k) - F_{n_k}(x_k) \leq F(\xi^-) - F_{n_k}(r_{m}^{a}) \\
  \leq F(\xi^-) - F(r_{m}^{a}) + F(r_{m}^{b}) - F_{n_k}(r_{m}^{a}).
  \]

- **Case 3:** \( \{x_k\} \) is decreasing and \( x_k \geq \xi \):
  \[
  \epsilon \leq F(x_k) - F_{n_k}(x_k) \leq F(r_{m}^{b}) - F(\xi) \\
  \leq F(r_{m}^{b}) - F(r_{m}^{a}) + F(r_{m}^{a}) - F_{n_k}(r_{m}^{a}) + F_{n_k}(\xi^-) - F_{n_k}(\xi).
  \]

- **Case 4:** \( \{x_k\} \) is decreasing and \( x_k \leq \xi \):
  \[
  \epsilon \leq F_{n_k}(x_k) - F(x_k) \leq F_{n_k}(r_{m}^{b}) - F(\xi) \\
  \leq F_{n_k}(r_{m}^{b}) - F_{n_k}(r_{m}^{a}) + F_{n_k}(r_{m}^{a}) - F(r_{m}^{a}) + F(r_{m}^{b}) - F(\xi).
  \]

In each case, for fixed \( k, m \) can be chosen such that \( r_{m}^{b} \) and \( r_{m}^{a} \) are arbitrarily close to \( \xi \). 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

\[
\tilde{G}_n(c_nS_0^n|X_n) = [\tilde{G}_n(c_nS_0^n|X_n(\omega)) - G(c_nS_0^n)] \\
+ [G(c_nS_0^n) - G_n(c_nS_0^n|X_n(\omega))] \\
+ G_n(c_nS_0^n|X_n).
\]

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

\[
\alpha_1 < \alpha < \alpha_2, \\
P[G_n(c_nS_0^n) \leq \alpha_1] \leq P[G_n(c_nS_0^n) \leq \alpha] \leq P[G_n(c_nS_0^n) \leq \alpha_2].
\]

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

\[
P[G_n(c_nS_0^n) \leq \alpha] = P\{G_n(c_nS_0^n) \leq \alpha_1 \cup \{\alpha_1 < G_n(c_nS_0^n) \leq \alpha\}\}
= P[G_n(c_nS_0^n) \leq \alpha_1] + 0,
\]

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

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

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,
\[ P[\tilde{G}_n(c_nS^0_0) \leq \alpha] = P[G_n(c_nS^0_0) \leq \alpha] + o_p(1) \leq \alpha + o_p(1), \]  
(B.31)
that is,
\[ \lim_{n \to \infty} P[\tilde{G}_n(c_nS^0_0) \leq \alpha] \leq \alpha, \]  
(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 \(U[0, 1]\) distribution, \(S^0_0\) the observed statistic and \(S^n(N) = (S^n_0, \ldots, S^n_N)\), a vector of \(N\) independent replications drawn from \(\tilde{G}_n(x)\). The randomized empirical survival function of \(S^0_0\) conditional on \(X\), under the null, is given by
\[ \tilde{G}^r_{N,n}[x, n, U_0, S^0_0, S^n(N), U(N)|X_n] = 1 - \frac{1}{N} \sum_{j=1}^{N} s(x - c_nS^n_j) + \frac{1}{N} \sum_{j=1}^{N} \delta(c_nS^n_j - x)s(U_j - U_0) \]  
(B.33)
with \(u(x) = 1_{[0,\infty)}(x), \delta(x) = 1_{\{0\}}\). The corresponding randomized empirical \(p\)-value is
\[ \tilde{p}^r_N(x) = \frac{N \tilde{G}^r_{N,n}(x) + 1}{N + 1}. \]  
(B.34)

Usually, validity of Monte Carlo testing is based on the fact the vector \((c_nS^0_0, \ldots, c_nS^n_N)\) 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 \((c_nS^0_0, \ldots, c_nS^n_N)\) 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 \((c_nS^0_0, \ldots, c_nS^n_N)\) is asymptotically exchangeable, we show that for any permutation \(\pi : [1, N] \to [1, N]\),
\[ \lim_{n \to \infty} P[S^0_0 \geq t_0, S^0_1 \geq t_1, \ldots, S^0_N \geq t_N] - P[S^0_{\pi(0)} \geq t_0, S^0_{\pi(1)} \geq t_1, \ldots, S^0_{\pi(N)} \geq t_N] = 0. \]

First, let rewrite
\[ P[S^0_0 \geq t_0, S^0_1 \geq t_1, \ldots, S^0_N \geq t_N] = E_{X_n}\{P[S^0_0 \geq t_0, S^0_1 \geq t_1, \ldots, S^0_N \geq t_N, X_n = x_n]\}. \]

Hence, if we use the conditional independence of the signs vectors (replicated and observed), we obtain
\[ P[S^0_0 \geq t_0, S^0_1 \geq t_1, \ldots, S^0_N \geq t_N, X_n = x_n] = P[X_n = x_n] \prod_{i=0}^{N} P[S^0_i \geq t_i | X_n = x_n] \]
As each survival function converges with probability one to $G(x)$, we finally obtain (B.35). Hence, it is straightforward to see that for $\pi : [1, N] \rightarrow [1, N]$, we asymptotically (on $n$) also have

$$P[S^n_0 \geq t_{\pi(0)}, S^n_1 \geq t_1, \ldots, S^n_N \geq t_N, X_n = x_n] \rightarrow \prod_{i=0}^{N} G(t_i) \text{ with probability one.}$$

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

$$\lim_{n \to \infty} P[S^n_0 \geq t_0, S^n_1 \geq t_1, \ldots, S^n_N \geq t_N] - P[S^n_{\pi(0)} \geq t_0, S^n_{\pi(1)} \geq t_1, \ldots, S^n_{\pi(N)} \geq t_N] = 0.$$ 

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 replications is such that $\alpha(N + 1)$, is an integer. Then,

$$\lim_{n \to \infty} \bar{p}_N(c_n, S^n_0) \leq \alpha,$$

which was to be proved.
References


