The Econometrics Journal

Econometrics Journal (2009), volume **12**, pp. S19–S49.
doi: 10.1111/j.1368-423X.2009.00285.x

Finite-sample distribution-free inference in linear median regressions under heteroscedasticity and non-linear dependence of unknown form

ELISE COUDIN[†] AND JEAN-MARIE DUFOUR[‡,§,¶]

[†] *Centre de Recherche en Économie et Statistique, Institut National de la Statistique et des Études Economiques, 15 Boulevard Gabriel Perl, 92245, Malakoff Cedex, France*
E-mail: `elise.coudin@ensae.fr`

[‡] *Department of Economics, McBill University, 855 Sherbrooke Street West, Montréal, Quebec H3A 2T7, Canada*

[§] *Centre interuniversitaire de recherche en analyse des organisations, 2020 rue University, 25e étage, Montréal, Quebec H3A 2A5, Canada*

[¶] *Centre interuniversitaire de recherche en économie quantitative, Université de Montréal, Quebec H3C 3J7, Canada*
E-mail: `jean-marie.dufour@mcgill.ca`

First version received: August 2008; final version accepted: January 2009

Summary We construct finite-sample distribution-free tests and confidence sets for the parameters of a linear median regression, where no parametric assumption is imposed on the noise distribution. The set-up studied allows for non-normality, heteroscedasticity, non-linear serial dependence of unknown forms as well as for discrete distributions. We consider a *mediangale* structure—the median-based analogue of a martingale difference—and show that the signs of mediangale sequences follow a nuisance-parameter-free distribution despite the presence of non-linear dependence and heterogeneity of unknown form. We point out that a simultaneous inference approach in conjunction with sign transformations yield statistics with the required pivotality features—in addition to usual robustness properties. Monte Carlo tests and projection techniques are then exploited to produce finite-sample tests and confidence sets. Further, under weaker assumptions, which allow for weakly exogenous regressors and a wide class of linear dependence schemes in the errors, we show that the procedures proposed remain asymptotically valid. The regularity assumptions used are notably less restrictive than those required by procedures based on least absolute deviations (LAD). Simulation results illustrate the performance of the procedures. Finally, the proposed methods are applied to tests of the drift in the Standard and Poor's composite price index series (allowing for conditional heteroscedasticity of unknown form).

Keywords: *Bootstrap*, *Discrete distribution*, *Distribution-free*, *GARCH*, *Heteroscedasticity*, *Median regression*, *Monte Carlo test*, *Non-normality*, *Projection methods*, *Quantile regression*, *Serial dependence*, *Signs*, *Sign test*, *Simultaneous inference*, *Stochastic volatility*.

1. INTRODUCTION

Median regression (and related quantile regressions) provides an attractive bridge between parametric and non-parametric models. Distributional assumptions on the disturbance process are relaxed, but the functional form remains parametric. Associated estimators, such as the least absolute deviations (LAD) estimator, are more robust to outliers than usual least-squares (LS) methods and may be more efficient whenever the median is a better measure of location than the mean (Dodge, 1997). They are especially appropriate when unobserved heterogeneity is suspected in the data. The current expansion of such ‘semiparametric’ techniques reflects an intention to depart from restrictive parametric frameworks (see Powell, 1994). However, related tests remain usually based on asymptotic normality approximations.

In this paper, we show that tests based on residual signs yield an entire system of finite-sample exact inference under very general assumptions. We study 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 set-up covers non-stochastic heteroscedasticity, standard conditional heteroscedasticity (like ARCH, GARCH, stochastic volatility models, . . .) as well as other forms of non-linear dependence. We provide both finite-sample and asymptotic distributional theories. In the first set of results, we show that the level of the tests is provably equal to the nominal level, for any sample size. Exact tests and confidence regions are valid under general assumptions and allow for heteroscedasticity and non-linear dependence of unknown forms, as well as for *discrete* distributions. This is done, in particular, by combining Monte Carlo tests adapted to discrete statistics—using a randomized tie-breaking procedure (Dufour, 2006)—with projection techniques, which allow inference on general parameter transformations (Dufour, 1990). We also show that the tests proposed include locally optimal tests. However, for more general processes that may involve stationary ARMA disturbances, sign-based statistics are no longer pivotal. The serial dependence parameters constitute nuisance parameters.

In a second set of results, we show that the proposed procedures remain asymptotically valid when the regressors are weakly exogenous and disturbances are stationary ARMA. Transforming sign-based statistics with standard heteroscedasticity and autocorrelation-corrected (HAC) methods allows one to eliminate nuisance parameters asymptotically. We thus extend the validity of the Monte Carlo test method. In such cases, we lose exactness but retain asymptotic validity. The latter holds under much weaker assumptions on moments or the shape of the distribution (such as the existence of a density) than usual asymptotically justified inference (such as LAD-based techniques). Besides, one does not need to evaluate the disturbance density at zero, which constitutes one of the major difficulties of asymptotic kernel-based methods associated with LAD and other quantile estimators.

A basic motivation for the sign-based techniques studied in this paper comes from an impossibility result due to Lehmann and Stein (1949), who proved that inference procedures that are valid under conditions of heteroscedasticity of unknown form when the number of observations is finite, must control the level of the tests conditional on the absolute values (see also Pratt and Gibbons, 1981). This result has two main consequences. First, sign-based methods constitute the only general way of producing provably valid inference for any given sample size. Second, all other methods, including the usual HAC methods developed by White (1980), Newey and West (1987), Andrews (1991) and others, which are not based on signs, are not provably valid for any sample size. Although this provides a compelling argument for using sign-based procedures, the latter have barely been exploited in econometrics; for a few exceptions which focus on simple time series models, see Dufour (1981), Campbell and Dufour (1991, 1995, 1997) and Wright (2000). In a regression context, the vast majority of the statistical literature is reviewed by Boldin et al. (1997). These authors also develop sign-based inference and estimation for linear models, both exact and asymptotic with i.i.d. errors. In the same vein, the recent paper by Chernozhukov et al. (2008) considers quantile regression models and derives finite sample inference using quantile indicators when the observations are independent.

The problem of interest in the present paper consists in giving conditions under which signs will be i.i.d. according to a known distribution, even though the variables to which indicator functions are applied are not independent or do not satisfy other regularity conditions (such as following an absolutely continuous distribution). An important feature of our results consists in allowing for a dynamic structure in the error distribution, providing a considerable extension of earlier results on the distribution of signs in the presence of dependent observations. Moreover, errors with discrete distribution (or mixtures of discrete and continuous distributions) are allowed, as opposed to the usual continuity assumption. This is made possible by the combination of a ternary sign operator—rather than binary—and Monte Carlo test techniques involving randomized tie-breaking.

Sign-based inference methods constitute an alternative to inference derived from the asymptotic distribution of LAD estimators and their extensions (see Koenker and Bassett, 1978, Powell, 1984, Weiss, 1991, Fitzenberger, 1997b, Horowitz, 1998, Zhao, 2001, etc.). An important problem in the LAD literature consists in providing good estimates of the asymptotic covariance matrix, on which inference relies. Powell (1984) suggested kernel estimation, but the most widespread method of estimation is the bootstrap (Buchinsky, 1995; Fitzenberger, 1997b; Hahn, 1997).[1] Kernel techniques are sensitive to the choice of kernel function and bandwidth parameter, and the estimation of the LAD asymptotic covariance matrix needs a reliable estimator of the error term density at zero. This may be tricky especially when disturbances are heteroscedastic or simply do not possess a density with respect to the Lebesgue measure (discrete distributions). Besides, whenever the normal distribution is not a good finite-sample approximation, inference based on covariance matrix estimation may be problematic. From a finite-sample point of view, asymptotically justified methods can be arbitrarily unreliable. Test sizes can be far from their nominal levels. One can find examples of such distortions for time series in Dufour (1981) and Campbell and Dufour (1995, 1997) and for L_1-estimation in Dielman and Pfaffenberger (1988a,b), De Angelis et al. (1993) and Buchinsky (1995). Inference based on signs constitutes an alternative that does not suffer from these shortcomings.[2]

The paper is organized as follows. In Section 2, we present the model and the notations. Section 3 contains results on exact inference. In Section 4, we derive confidence intervals at any given confidence level and illustrate the method on a numerical example. Section 5 is dedicated to the asymptotic validity of the finite-sample inference method. In Section 6, we give simulation results from comparisons with usual techniques. Section 7 presents an illustrative application: testing the presence of a drift in the Standard and Poor's composite price index series. Section 8 concludes. The Appendix contains the proofs.

[1] See Buchinsky (1995, 1998) for a review and Fitzenberger (1997b) for a comparison between these methods.

[2] Other notable areas of investigation in the L_1-literature concern: (1) censored quantile regressions (Powell, 1984, 1986, Fitzenberger, 1997a, Buchinsky and Hahn, 1998), (2) endogeneity (Amemiya, 1982, Powell, 1983, Hong and Tamer, 2003), (3) misspecification (Jung, 1996, Kim and White, 2002, Komunjer, 2005).

2. FRAMEWORK

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

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

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

Inference on this model will be made possible through assumptions on the conditional medians of the errors. To do this, it will be convenient to consider *adapted* sequences of the form

$$\mathcal{S}(\mathbf{v},\ \mathcal{F}) = \{v_t, \mathcal{F}_t : t = 1, 2,\ \ldots\}, \tag{2.2}$$

where v_t is any measurable function of $W_t = (y_t,\ x_t')'$, $\mathcal{F}_t$ is a σ-field in Ω, $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s < t$, $\sigma(W_1,\ \ldots,\ W_t) \subset \mathcal{F}_t$ and $\sigma(W_1, \ldots, W_t)$ is the σ-algebra spanned by $W_1, \ldots, W_t$.

We shall depart from the usual assumption that $E(u_t|\mathcal{F}_{t-1}) = 0,\ \forall t \geq 1$, i.e. $\mathbf{u} = \{u_t : t = 1, 2, \ldots\}$ in the adapted sequence $\mathcal{S}(\mathbf{u},\ \mathcal{F}) = \{u_t, \mathcal{F}_t : t = 1, 2,\ \ldots\}$ is a martingale difference with respect to $\mathcal{F}_t = \sigma(W_1,\ \ldots,\ W_t),\ t = 1, 2,\ \ldots$.

In a framework that allows for heteroscedasticity of unknown form, it is known from Bahadur and Savage (1956) that inference on the mean of i.i.d. observations of a random variable, without any further assumption on the form of the 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. The median is especially appropriate if the distribution of the disturbance process does not possess moments. Thus, in the median regression framework, it appears that the martingale difference assumption should be replaced by an analogue in terms of median. Such a *mediangale* may be defined conditional on the design matrix X or unconditionally. Here, we focus on the conditional form.

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

The above definition allows u_t to have a discrete distribution with a non-zero probability mass at zero. A more restrictive version, called the strict conditional mediangale, imposes a zero probability mass at zero. Then, $\mathsf{P}[u_1 < 0|X] = \mathsf{P}[u_1 > 0|X] = 0.5$ and $\mathsf{P}[u_t < 0|u_1,\ \ldots,\ u_{t-1}, X] = \mathsf{P}[u_t > 0|u_1,\ \ldots,\ u_{t-1}, X] = 0.5$, for $t > 1$. With no mass at zero and no matrix X, this concept coincides with the mediangale one defined in Linton and Whang (2007), together with other quantilegales.[3]

[3] Linton and Whang (2007) define that u_t is a mediangale if $E(\psi_{\frac{1}{2}}(u_t)|\mathcal{F}_{t-1}) = 0,\ \forall t$, where $\mathcal{F}_{t-1} = \sigma(u_{t-1}, u_{t-2},\ \ldots)$ and $\psi_{\frac{1}{2}}(x) = \frac{1}{2} - \mathbf{1}_{(-\infty,0)}(x)$. This definition is adapted to continuous distributions but does not work

Stating that $\mathbf{u}$ is a weak mediangale with respect to $\mathcal{F}$ is equivalent to assuming that its sign process $\mathbf{s(u)} = \{s(u_t) : t = 1, 2, \dots\}$, where $s(a) = \mathbf{1}_{[0,+\infty)}(a) - \mathbf{1}_{(-\infty,0]}(a)$, $\forall a \in \mathbb{R}$, is a martingale difference with respect to the same sequence of sub-σ algebras $\mathcal{F}$. The difference of martingale assumption on the raw process $\mathbf{u}$ is replaced by a quasi-similar hypothesis on a robust transform of this process $\mathbf{s(u)}$.

However, the weak conditional mediangale concept differs from a martingale difference on the signs, because it requires conditioning upon the whole process X. We shall see later that asymptotic inference may be available under a classical martingale difference on signs or, more generally, mixing conditions on $\{s(u_t), \sigma(W_1, \dots, W_t) : t = 1, 2, \dots\}$.

It is relatively easy to deal with a weak mediangale by a simple transformation of the sign operator. Consider $\mathsf{P}[u_t = 0 \mid X,\ u_1,\ \dots,\ u_{t-1}] = p_t(X,\ u_1,\ \dots,\ u_{t-1}) > 0$, where the $p_t(\cdot)$ are unknown and may vary between observations. A way out consists in modifying the sign function $s(x)$ as $\tilde{s}(x,\ V) = s(x) + [1 - s(x)^2]s(V - 0.5)$, where $V \sim \mathcal{U}(0, 1)$. If V_t is independent of u_t then, irrespective of the distribution of u_t,

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

To simplify the presentation, we shall focus on the strict mediangale concept. Therefore, our model will rely on the following assumption.

ASSUMPTION 2.1. (*Strict conditional mediangale*). *The components of* $u = (u_1, \dots, u_n)'$ *satisfy a strict mediangale conditional on X.*

One remark concerns exogeneity. As long as the x_t's are strongly exogenous, the conditional mediangale concept is equivalent to a martingale difference on signs with respect to $\mathcal{F}_t = \sigma(W_1,\ \dots,\ W_t),\ \ t = 1,\ 2,\ \dots.$

PROPOSITION 2.1. (*Mediangale exogeneity*). *Suppose* $\{x_t\colon\ t = 1, 2, \dots\}$ *is a strongly exogenous process for* β, $\mathsf{P}[u_1 > 0] = \mathsf{P}[u_1 < 0] = 0.5$ *and*

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

Then $\{u_t : t = 1, 2, \dots\}$ *is a strict mediangale conditional on X.*

Model (2.1) with the Assumption 2.1 allows for very general forms of the disturbance distribution, including asymmetric, heteroscedastic or dependent ones, as long as conditional medians are 0. Neither density nor moment existence are required. Indeed, what the mediangale concept requires is a form of independence in the signs of the residuals. This extends results in Dufour (1981), Campbell and Dufour (1991, 1995, 1997) and Dufour et al. (1998).

For example, Assumption 2.1 is satisfied if $u_t = \sigma_t(x_1, \dots, x_n)\varepsilon_t, t = 1, \dots, n$, where $\varepsilon_1, \dots, \varepsilon_n$ are i.i.d. conditional on X, which is relevant for cross-sectional data. Many dependence schemes are also covered, especially any model of the form $u_1 = \sigma_1(x_1, \dots, x_{t-1})\varepsilon_1, u_t = \sigma_t(x_1, \dots, x_{t-1}, u_1, \dots, u_{t-1})\varepsilon_t$, $t = 2, \dots, n$, where $\varepsilon_1, \dots, \varepsilon_n$ are independent with median $0, \sigma_1(x_1, \dots, x_{t-1})$ and $\sigma_t(x_1, \dots, x_n, u_1, \dots, u_{t-1}), t = 2, \dots, n$, are non-zero with probability one. In time series context, this includes models presenting robustness properties to endogenous disturbance variance (or volatility) specification, such as ARCH, GARCH or stochastic volatility

well with discrete distributions. If u_t has a mass at zero, the condition given by Definition 2.1 can hold even if $\mathsf{E}(\psi_{\frac{1}{2}}(u_t)|\mathcal{F}_{t-1}) \neq 0$.

models with non-Gaussian noises. Further, the mediangale property is more general because it does not specify explicitly the functional form of the variance in contrast with an ARCH specification. Note again that the disturbance process does not have to be second-order stationary.

Asymptotic normality of the LAD estimator, which is presented in its most general way in Fitzenberger (1997b), holds under some mixing concepts on $\{s(u_t), \sigma(W_1, \ldots, W_t) : t = 1, 2, \ldots\}$ and an orthogonality condition between $s(u_t)$ and x_t. Besides, it requires additional assumptions on moments.[4] With such a choice, testing is necessarily based on approximations (asymptotic or bootstrap). Here, we focus on valid finite-sample inference without any further assumption on the form of the distributions. This non-parametric set-up extends those used in Dufour (1981) and Campbell and Dufour (1991, 1995, 1997).

Assumption 2.1 can easily be extended to allow for another quantile q by setting $\mathsf{P}[u_t < 0|\mathcal{F}_{t-1}] = q, \ \forall t$, which would lead to $\mathsf{P}[u_t < 0|u_1, \ldots, u_{t-1}, x_1, \ldots, x_t] = q$ in Proposition 2.1. However, with error heterogeneity or dependence of unknown form, such an assumption can plausibly hold only for a single quantile. So little generality is lost by focusing on the median case. Further, contrary to other quantiles, the median may have an economic meaning when it coincides with the expectation, e.g. if the error density is symmetric. It can be used to state expectation-based economic conditions such as a no-arbitrage opportunity condition on a market etc.

A classical result in non-parametric statistics consists in using this Bernoulli distribution to build exact tests and confidence intervals on quantiles (for i.i.d. observations); see Thompson (1936), Scheffé and Tukey (1945) and the review of David (1981, ch. 2). For recent econometric exploitation of a quantile version of this result which holds if the observations are X-conditionally independent, see Chernozhukov et al. (2008). Proposition 2.1 above provides general conditions under which such a result holds for non-i.i.d. observations. Finally, the set-up presented here extends those approaches to the time series context where some kinds of Markovian serial dependence are permitted as well as discrete distributions.

3. EXACT FINITE-SAMPLE SIGN-BASED INFERENCE

In finite samples, first-order asymptotic approximations can be misleading. Test sizes of asymptotically justified t- or χ^2-statistics can be quite far from their nominal level. One can find examples of such distortions in the dynamic literature (see, for example, Dufour, 1981, Mankiw and Shapiro, 1986, Campbell and Dufour, 1995, 1997); on inference based on L_1-estimators (see Dielman and Pfaffenberger, 1988a,b; Buchinsky, 1995; De Angelis et al., 1993). 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 is satisfied in finite samples. The asymptotic method unreliability motivates us to turn a fully finite-sample-based approach. Sign-based procedures provide a way to build distribution-free statistics even in finite samples. Sign-based statistics have been used in the statistical literature to derive non-parametric sign tests.

In this section, we present the general sign pivotality result and apply it in median regression context to derive sign-based test statistics that are pivots and provide power against alternatives

[4] Fitzenberger (1997b) show that LAD and quantile estimators are consistent and asymptotically normal when $\mathsf{E}[x_t s_\theta(u_t)] = 0, \ \forall t$, where (u_t, x_t) has a density and finite second moments.

of interest. This will enable us to build Monte Carlo tests relying on their exact distribution. Therefore, the level of those tests is exactly controlled for any sample size. We study first the test problem, then build confidence sets. Finally, estimators can be derived.[5] Hence, results on the valid finite-sample test problem will be adapted to obtain valid confidence intervals and estimators.

3.1. Distribution-free pivotal functions and non-parametric tests

When the disturbance process is a conditional mediangale, the joint distribution of the signs of the disturbances is completely determined. If there is no positive mass at zero, the signs are i.i.d. and take the values 1 and −1 with equal probability 1/2. The case with a mass at zero can be covered provided the transformation in the sign operator definition presented in the previous section. These results are stated more precisely in the following propositions.

PROPOSITION 3.1. (*Sign distribution*). *Under model (2.1), suppose the errors* $(u_1, \ldots, u_n)$ *satisfy a strict mediangale conditional on* $X = [x_1, \ldots, x_n]'$. *Then the variables* $s(u_1), \ldots, s(u_n)$ *are i.i.d. conditional on X according to the distribution*

$$\mathsf{P}[s(u_t) = 1\,|x_1, \ldots, x_n] = \mathsf{P}[s(u_t) = -1\,|x_1, \ldots, x_n] = \frac{1}{2}, \quad t = 1, \ldots, n. \tag{3.1}$$

More generally, this result holds for any combination of $t = 1, \ldots, n$. If there is a permutation $\pi : i \to j$ such that mediangale property holds for *j*, then the signs are i.i.d. From Proposition 3.1, it follows that the residual sign vector

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

has a nuisance-parameter-free distribution (conditional on *X*), i.e. it is a ‘pivotal function’. Its distribution is easy to simulate from a combination of *n* independent uniform Bernoulli variables. Furthermore, any function of the form $T = T(s(y - X\beta), X)$ is pivotal, conditional on *X*. Once the form of *T* is specified, the distribution of the statistic *T* is totally determined and can also be simulated.

Using Proposition 3.1, it is possible to construct tests for which the size is fully controlled in finite samples. Consider testing $H_0(\beta_0) : \beta = \beta_0$ against $H_1(\beta_0) : \beta \neq \beta_0$. Under $H_0(\beta_0)$, $s(y_t - x'_t\beta_0) = s(u_t), t = 1, \ldots, n$. Thus, conditional on *X*,

$$T\big(s(y - X\beta_0), X\big) \sim T(S_n, X), \tag{3.2}$$

where $S_n = (s_1, \ldots, s_n)$ and $s_1, \ldots, s_n \overset{\text{i.i.d.}}{\sim} \mathcal{B}(1/2)$. A test with level α rejects $H_0(\beta_0)$ when

$$T\big(s(y - X\beta_0), X\big) > c_T(X, \alpha), \tag{3.3}$$

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

This result is generalized for distributions with a positive mass at zero in the following proposition.

PROPOSITION 3.2. (*Randomized sign distribution*). *Suppose (2.1) holds with the assumption that* $u_1, \ldots, u_n$ *belong to a weak mediangale conditional on X. Let* $V_1, \ldots, V_n$ *be i.i.d. random variables* $\mathcal{U}(0, 1)$ *distributed and independent of* $u_1, \ldots, u_n$ *and X. Then the variables* $\tilde{s}_t =$

[5] For the estimation theory, the reader is referred to Coudin and Dufour (2006).

$\tilde{s}(u_t,\ V_t)$ are i.i.d. conditional on X with the distribution $\mathsf{P}[\tilde{s}_t = 1 \mid X] = \mathsf{P}[\tilde{s}_t = -1 \mid X] = \frac{1}{2}$, $t = 1, \ldots, n$.

All the procedures described in the paper can be applied by replacing s by $\tilde{s}$. When the error distributions possess a mass at zero, the test statistic $T(\tilde{s}(y - X\beta_0), X)$ has to be used instead of $T(s(y - X\beta_0), X)$.

3.2. Regression sign-based statistics

We consider test statistics of the following form:

$$D_S(\beta_0, \Omega_n) = s(y - X\beta_0)' X \Omega_n\big(s(y - X\beta_0), X\big) X' s(y - X\beta_0), \tag{3.4}$$

where $\Omega_n(s(y - X\beta_0), X)$ is a $p \times p$ weight matrix that depends on the *constrained* signs $s(y - X\beta_0)$ under $H_0(\beta_0)$. The weight matrix $\Omega_n(s(y - X\beta_0), X)$ provides a standardization that can be useful for power considerations as well as to account for dependence schemes that cannot be eliminated by the sign transformation. Further, $\Omega_n(s(y - X\beta_0), X)$ would normally be selected to be positive definite (although this is not essential to show the pivotality of the test statistic under the null hypothesis).[6]

Statistics of the form $D_S(\beta_0, \Omega_n)$ include as special cases the ones studied by Koenker and Bassett (1982) and Boldin et al. (1997). Namely, on taking $\Omega_n = I_p$ and $\Omega_n = (X'X)^{-1}$, we get:

$$SB(\beta_0) = s(y - X\beta_0)' X X' s(y - X\beta_0) = \left\| X' s(y - X\beta_0) \right\|^2, \tag{3.5}$$

$$SF(\beta_0) = s(y - X\beta_0)' P(X) s(y - X\beta_0) = \left\| X' s(y - X\beta_0) \right\|_M^2, \tag{3.6}$$

where $P(X) = X(X'X)^{-1}X'$. In Boldin et al. (1997), it is shown that $SB(\beta_0)$ and $SF(\beta_0)$ can be associated with locally most powerful tests in the case of i.i.d. disturbances under some regularity conditions on the distribution function (especially $f'(0) = 0$).[7] Their proof can easily be extended to disturbances that satisfy the mediangale property and for which the conditional density at zero is the same $f_t(0|X) = f(0|X), t = 1, \ldots, n$.

$SF(\beta_0)$ can be interpreted as a sign analogue of the Fisher statistic. $SF(\beta_0)$ is a monotonic transformation of the Fisher statistic for testing $\gamma = 0$ in the regression of $s(y - X\beta_0)$ on $X : s(y - X\beta_0) = X\gamma + v$. This remark holds also for a general sign-based statistic of the form (3.6), when $s(y - X\beta_0)$ is regressed on $\Omega_n^{-1/2} X$.

Wald, Lagrange multiplier (LM) and likelihood ratio (LR) asymptotic tests for M-estimators, such as the LAD estimator, in L_1-regression are developed by Koenker and Bassett (1982). They

[6] Under more restrictive assumptions, statistics that exploit other robust functions of $y - X\beta_0$ (such as ranks, signed ranks, and signs and ranks) can lead to more powerful tests. However, the fact we allow for both heteroscedasticity and non-linear serial dependence of unknown forms appears to break the required pivotality result and makes the use of such statistics quite difficult if not impossible in the context of our set-up. For discussion of such alternative statistics (applicable under stronger assumptions), see Hallin and Puri (1991, 1992), Hallin et al. (2006, 2008), Hallin and Werker (2003) and the references therein.

[7] The power function of the locally most powerful sign-based test has the faster increase when departing from β_0. In the multiparameter case, the scalar measure required to evaluate that speed is the curvature of the power function. Restricting to unbiased tests, Boldin et al. (1997) introduced different locally most powerful tests corresponding to different definitions of curvature. $SB(\beta_0)$ maximizes the mean curvature, which is proportional to the trace of the shape; see Dubrovin et al. (1984, ch. 2, pp. 76–86) or Gray (1998, ch. 21, pp. 373–80) for a discussion of various curvature notions.

assume i.i.d. errors and a fixed design matrix. In that set-up, the LM statistic for testing $H_0(\beta_0): \beta = \beta_0$ turns out to be the $SF(\beta_0)$ statistic. The same authors also remarked that this type of statistic is asymptotically nuisance-parameter-free, contrary to LR and Wald-type statistics.

The Boldin et al. (1997) local optimality interpretation can be extended to heteroscedastic disturbances. In such a case, the locally optimal test statistic associated with the mean curvature, i.e. the test with the highest power near the null hypothesis according to a trace argument, will be of the following form.

PROPOSITION 3.3. *In model (2.1), suppose the mediangale Assumption 2.1 holds, and the disturbances are heteroscedastic with conditional densities $f_t(\cdot\,|X), t = 1, 2, \ldots,$ which are continuously differentiable around zero and such that $f_t'(0|X) = 0$. Then, the locally optimal sign-based statistic associated with the mean curvature is*

$$\tilde{SB}(\beta_0) = s(y - X\beta_0)'\tilde{X}\tilde{X}'s(y - X\beta_0), \tag{3.7}$$

where $\tilde{X} = \text{diag}(f_1(0|X), \ldots, f_n(0|X))X$.

When the $f_i(0|x)$'s are unknown, the optimal statistic is not feasible. The optimal weights must be replaced by approximations, such as weights derived from the normal distribution.

Sign-based statistics of the form (3.4) can also be interpreted as GMM statistics which exploit the property that $\{s_t \otimes x_t', \mathcal{F}_t\}$ is a martingale difference sequence.[8] However, these are quite unusual GMM statistics. Indeed, the parameter of interest is not defined by moment conditions in explicit form. It is implicitly defined as the solution of some robust estimating equations (involving constrained signs):

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

For i.i.d. disturbances, Godambe (2001) showed that these estimating functions are optimal among all the linear unbiased (for the median) estimating functions $\sum_{t=1}^{n} a_t(\beta)s(y_t - x_t'\beta)$. For independent heteroscedastic 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 (resp. $\tilde{X}$) can be viewed as optimal instruments for the linear model.

We now turn to linearly 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 it by $J_n(s(y - X\beta_0), X)$. We consider $\Omega_n(s(y - X\beta_0), X) = \frac{1}{n}\hat{J}_n(s(y - X\beta_0), X)^{-1}$, where $\hat{J}_n(s(y - X\beta_0), X)$ stands for a consistent estimate of $J_n(s(y - X\beta_0), X)$ that can be obtained using kernel estimators; for example, see Parzen (1957), Newey and West (1987), Andrews (1991) and White (2001). This leads to

$$D_S\left(\beta_0, \frac{1}{n}\hat{J}_n^{-1}\right) = \frac{1}{n}s(y - X\beta_0)'X\hat{J}_n^{-1}X's(y - X\beta_0). \tag{3.8}$$

$J_n(s(y - X\beta_0), X)$ accounts for dependence among signs and explanatory variables. Hence, by using an estimate of its inverse as weighting matrix, we perform a HAC correction. Note that the correction depends on β_0.

[8] Concerning power performance again, Chernozhukov et al. (2008) show also the class of GMM sign-based statistics contains a locally asymptotically uniformly most powerful invariant test.

In all cases, $H_0(\beta_0)$ is rejected when the statistic evaluated at $\beta = \beta_0$ is large: $D_S(\beta_0, \Omega_n) > c_{\Omega_n}(X, \alpha)$, where $c_{\Omega_n}(X, \alpha)$ is a critical value, which depends on the level α. Since we are looking at pivotal functions, the critical values can be evaluated to any degree of precision by simulation. This is the strategy followed by Chernozhukov et al. (2008), which exploits the same finite sample property of (θ-) signs in a quantile regression context with conditionally independent observations. However, as the distribution is discrete, a test based on $c_{\Omega_n}(X, \alpha)$ may not exactly reach the nominal level. A more elegant solution consists in using the technique of *Monte Carlo tests* with a randomized tie-breaking procedure, which do not suffer from this shortcoming. Further, we will show later that the Monte Carlo procedure also enables one to build tests with asymptotically controlled level for general processes when Assumption 2.1 fails to hold.

3.3. *Monte Carlo tests*

Monte Carlo tests can be viewed as a finite-sample version of the bootstrap. They have been introduced by Dwass (1957) (see also Barnard, 1963) and can be adapted to any pivotal statistic whose distribution can be simulated. For a general review and for extensions in the case of the presence of a nuisance parameter, the reader is referred to Dufour (2006).

In the case of *discrete distributions*, the method must be adapted to deal with ties. Here, we use a randomized tie-breaking procedure for evaluating empirical survival functions (see Dufour, 2006). Let us consider a statistic T, whose conditional distribution given X is discrete and free of nuisance parameters, and a test which rejects the null hypothesis when $T \geq c(\alpha)$. Let $T^{(0)}$ be the observed value of T, and $T^{(1)}, \ldots, T^{(N)}$, N independent replicates of T. Each replication $T^{(j)}$ is associated with a uniform random variable $W^{(j)} \sim \mathcal{U}(0, 1)$ to produce the pairs $(T^{(j)}, W^{(j)})$. The vector $(W^{(0)}, \ldots, W^{(N)})$ is independent of $(T^{(0)}, \ldots, T^{(N)})$. $(T^{(i)}, W^{(i)})$'s are ordered according to

$$(T^{(i)},\ W^{(i)}) \geq (T^{(j)},\ W^{(j)}) \Leftrightarrow \{T^{(i)} > T^{(j)} \text{ or } (T^{(i)} = T^{(j)} \text{ and } W^{(i)} \geq W^{(j)})\}.$$

This leads to the following p-value function:

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

where the empirical survival function, $\tilde{G}_N(x) = 1 - \frac{1}{N}\sum_{i=1}^{N} s_+(x - T^{(i)}) + \frac{1}{N}\sum_{i=1}^{N} \delta(T^{(i)} - x)s_+(W^{(i)} - W^{(0)})$, with $s_+(x) = \mathbf{1}_{[0,\infty)}(x)$, $\delta(x) = \mathbf{1}_{\{0\}}$. Then

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

The randomized tie-breaking allows one to exactly control the level of the procedure. This may also increase the power of the test.

4. REGRESSION SIGN-BASED CONFIDENCE SETS

In this section, we discuss how to use Monte Carlo sign-based joint tests in order to build confidence sets for β with known level. This can be done as follows. For each value $\beta_0 \in \mathbb{R}^p$, perform the Monte Carlo sign test for $H_0(\beta_0)$ and get the associated simulated p-value. The confidence set $C_{1-\alpha}(\beta)$ that contains any β_0 with p-value higher than α has, by construction, level $1 - \alpha$ (see Dufour, 2006). From this simultaneous confidence set for β, it is possible, by *projection techniques*, to derive confidence intervals for the individual components. More generally, we can obtain conservative confidence sets for any transformation $g(\beta)$, where g can be any kind of real functions, including non-linear ones. Obviously, obtaining a continuous grid of $\mathbb{R}^p$ is not realistic. We will instead require *global optimization search algorithms*.

4.1. Confidence sets and conservative confidence intervals

Projection techniques yield finite-sample valid confidence intervals and confidence sets for general functions of the parameter β. For examples of use in different settings and for further discussion, the reader is referred to Dufour (1990, 1997), Abdelkhalek and Dufour (1998), Dufour and Kiviet (1998), Dufour and Jasiak (2001) and Dufour and Taamouti (2005). The basic idea is the following one. Suppose a simultaneous confidence set with level $1-\alpha$ for $\beta, C_{1-\alpha}(\beta)$, is available. Since $\beta \in C_{1-\alpha}(\beta) \implies g(\beta) \in g(C_{1-\alpha}(\beta))$, we have $\mathsf{P}[\beta \in C_{1-\alpha}(\beta)] \geq 1-\alpha \implies \mathsf{P}[g(\beta) \in g(C_{1-\alpha}(\beta))] \geq 1-\alpha$. Thus, $g(C_{1-\alpha}(\beta))$ is a conservative confidence set for $g(\beta)$. If $g(\beta)$ is scalar, the interval (in the extended real numbers) $I_g[C_{1-\alpha}(\beta)] = [\inf_{\beta \in C_{1-\alpha}(\beta)} g(\beta)\, , \sup_{\beta \in C_{1-\alpha}(\beta)} g(\beta)]$ has level $1-\alpha$:

$$\mathsf{P}\left[\inf_{\beta \in C_{1-\alpha}(\beta)} g(\beta) \leq g(\beta) \leq \sup_{\beta \in C_{1-\alpha}(\beta)} g(\beta)\right] \geqslant 1-\alpha.$$

Hence, to obtain valid conservative confidence intervals for the individual component β_k in the model (2.1) under mediangale Assumption 2.1, it is sufficient to solve the following numerical optimization problems, where s.c. stands for 'subject to the constraint':

$$\min_{\beta \in \mathbb{R}^p} \beta_k \quad \text{s.c. } \tilde{p}_N\big(D_S(\beta)\big) \geq \alpha, \qquad \max_{\beta \in \mathbb{R}^p} \beta_k \quad \text{s.c. } \tilde{p}_N\big(D_S(\beta)\big) \geq \alpha,$$

where $\tilde{p}_N$ is computed using N replicates $D_S^{(j)}$ of the statistic D_S under the null hypothesis. In practice, we use *simulated annealing* as optimization algorithm (see Goffe et al., 1994; Press et al., 1996).[9]

In the case of multiple tests, projection techniques allow to perform tests on an arbitrary number of hypotheses, without ever losing control of the overall level: rejecting at least one true null hypothesis will not exceed the specified level α.

4.2. Numerical illustration

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

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

We conduct an exact inference procedure with $N = 999$ replicates. The true process is generated with $\beta_0 = \beta_1 = 0$. We perform tests of $H_0(\beta^*) : \beta = \beta^*$ on a grid for $\beta^* = (\beta_0^*, \beta_1^*)$ and retain the associated simulated p-values. As β is a two-vector, we can provide a graphical illustration. To each value of the vector β is associated the corresponding simulated p-value. Confidence

[9] See Chernozhukov et al. (2008) for the use of other MCMC algorithms.

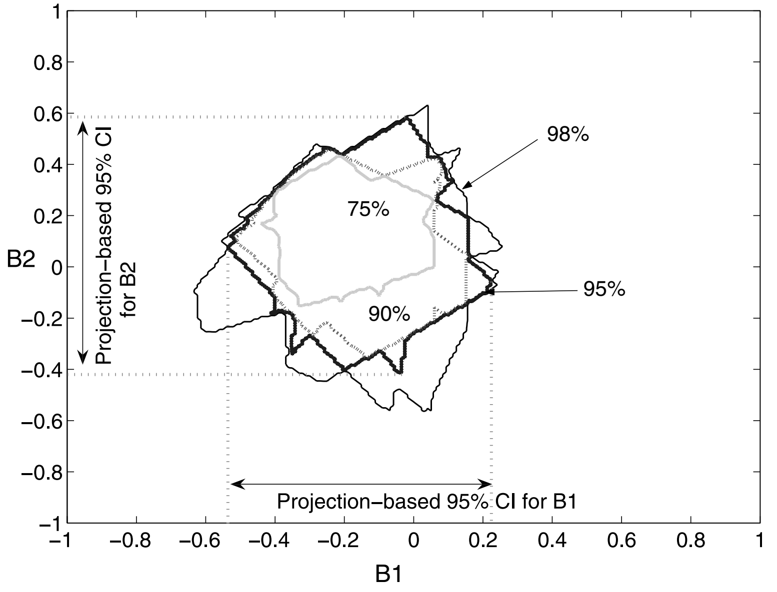


Figure 1. Confidence regions provided by SF-based inference.

Table 1. Confidence intervals.

		OLS	White	SF
β_0	95%CI	[−4.57, 0.82]	[−4.47, 0.72]	[−0.54, 0.23]
	98%CI	[−5.10, 1.35]	[−4.98, 1.23]	[−0.64, 0.26]
β_1	95%CI	[−2.50, 3.22]	[−1.34, 2.06]	[−0.42, 0.59]
	98%CI	[−3.07, 3.78]	[−1.67, 2.39]	[−0.57, 0.64]

region with level $1 - \alpha$ contains all the values of β with p-values greater than α. Confidence intervals are obtained by projecting the simultaneous confidence region on the axis of β_0 or β_1; see Figure 1 and Table 1.

The confidence regions so obtained increase with the level and cover other confidence regions with smaller level. Confidence regions are highly non-elliptic and thus may lead to different results than an asymptotic inference. Concerning confidence intervals, sign-based ones appear to be largely more robust than OLS and White CI and are less sensitive to outliers.

5. ASYMPTOTIC THEORY

This section is dedicated to asymptotic results. We point out that the mediangale Assumption 2.1 excludes some common processes, whereas usual asymptotic inference still can be conducted on them. We relax Assumption 2.1 to allow random X that may not be independent of u. We show that the finite-sample sign-based inference remains asymptotically valid. For a fixed number of replicates, when the number of observations goes to infinity, the level of a test tends to the nominal level. Besides, we stress the ability of our methods to cover heavy-tailed distributions, including infinite disturbance variance.

5.1. Asymptotic distributions of test statistics

In this part, we derive asymptotic distributions of the sign-based statistics. We show that the HAC-corrected version of the sign-based statistic $D_S(\beta_0, \frac{1}{n}\hat{J}_n^{-1})$ in (3.8) allows one to obtain an asymptotically pivotal function. The set of assumptions we make to stabilize the asymptotic behaviour will be needed for further asymptotic results. We consider the linear model (2.1), with the following assumptions:

ASSUMPTION 5.1. (*Mixing*). $\{(x'_t, u_t) : t = 1, 2, \dots\}$ *is* α*-mixing of size* $-r/(r-2), r > 2$.[10]

ASSUMPTION 5.2. (*Moment condition*). $E[s(u_t)x_t] = 0,\ t = 1, \dots, n,\ \forall n \in \mathbb{N}$.

ASSUMPTION 5.3. (*Boundedness*). $x_t = (x_{1t}, \dots, x_{pt})'$ *and* $E[|x_{ht}|^r] < \Delta < \infty,\ h = 1, \dots, p,\ t = 1, \dots, n,\ \forall n \in \mathbb{N}$.

ASSUMPTION 5.4. (*Non-singularity*). $J_n = \operatorname{var}[\frac{1}{\sqrt{n}} \sum_{t=1}^{n} s(u_t)x_t]$ *is uniformly positive definite.*

ASSUMPTION 5.5. (*Consistent estimator of* J_n). $\Omega_n(\beta_0)$ *is symmetric positive definite uniformly over n and* $\Omega_n - \frac{1}{n} J_n^{-1} \xrightarrow{p} 0$.

We can now give the following result on the asymptotic distribution of $D_S(\beta_0, \Omega_n)$ under $H_0(\beta_0)$.

THEOREM 5.1. (*Asymptotic distribution of sign-based statistics*). *In model (2.1), with Assumptions 5.1–5.5, we have, under* $H_0(\beta_0)$, $D_S(\beta_0, \Omega_n) \to \chi^2(p)$.

In particular, when the mediangale condition holds, J_n reduces to $E(X'X/n)$, and $(X'X/n)^{-1}$ is a consistent estimator of J_n^{-1}. This yields the following corollary.

COROLLARY 5.1. *In model (2.1), suppose the mediangale Assumption 2.1 and boundedness Assumption 5.3 are fulfilled. If* $X'X/n$ *is positive definite uniformly over n and converges in probability to a definite positive matrix, then, under* $H_0(\beta_0)$, $SF(\beta_0) \to \chi^2(p)$.

5.2. 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 methods.

5.2.1. Generalities. Let us consider a parametric or semi-parametric model $\{M_\beta,\ \beta \in \Theta\}$. Let $S_n(\beta_0)$ be a test statistic for $H_0(\beta_0)$. Let c_n be the rate of convergence. Under $H_0(\beta_0)$, the distribution function of $c_n S_n(\beta_0)$ is denoted by $F_n(x)$. We suppose that $F_n(x)$ converges almost everywhere to a distribution function $F(x)$. $G(x)$ and $G_n(x)$ are the corresponding survival functions. In Theorem 5.2, we show that if a sequence of conditional survival functions $\tilde{G}_n(x|X_n(\omega))$ given $X(\omega)$, satisfies $\tilde{G}_n(x|X_n(\omega)) \to G(x)$ with probability one, where G does not

[10] See White (2001) for a definition of α-mixing.

depend on the realization $X(\omega)$, then $\tilde{G}_n(x|X_n(\omega))$ can be used as an approximation of $G_n(x)$. It can be seen as a pseudo survival function of $c_n S_n(\beta_0)$.

THEOREM 5.2. (*Generic asymptotic validity*). *Let $S_n(\beta_0)$ be a test statistic for testing $H_0(\beta_0)$: $\beta = \beta_0$ against $H_1(\beta_0): \beta \neq \beta_0$ in model (2.1). Suppose that, under $H_0(\beta_0)$,*

$$\mathsf{P}[c_n S_n(\beta_0) \geq x|X_n] = G_n(x|X_n) = 1 - F_n(x|X_n) \underset{n\to\infty}{\to} G(x) \; a.e.,$$

where $\{c_n\}$ is a sequence of positive constants, and suppose that $\tilde{G}_n(x|X_n(\omega))$ is a sequence of survival functions such that $\tilde{G}_n(x|X_n(\omega)) \underset{n\to\infty}{\to} G(x)$ with probability one. Then

$$\lim_{n\to\infty} \mathsf{P}[\tilde{G}_n(c_n S_n(\beta_0), X_n(\omega)) \leq \alpha] \leq \alpha. \tag{5.1}$$

This theorem can also be stated in a Monte Carlo version. Following Dufour (2006), we use empirical survival functions and empirical p-values adapted to discrete statistics in a randomized way, but the replicates are not drawn from the same distribution as the observed statistic. However, both distribution functions, respectively F_n and $\tilde{F}_n$, converge to the same limit F. Let $U(N+1) = (U^{(0)}, U^{(1)}, \ldots, U^{(N)})$ be a vector of $N+1$ i.i.d. real variables drawn from a $\mathcal{U}(0, 1)$ distribution, $S_n^{(0)}$ is the observed statistic and $S_n(N) = (S_n^{(1)}, \ldots, S_n^{(N)})$ a vector of N independent replicates drawn from $\tilde{F}_n$. Then, the randomized pseudo empirical survival function under $H_0(\beta_0)$ is

$$\tilde{G}_n^{(N)}\big(x, n, S_n^{(0)}, S_n(N), U(N+1)\big) = 1 - \frac{1}{N}\sum_{j=1}^{N} s_+\big(x - c_n S_n^{(j)}\big) + \frac{1}{N}\sum_{j=1}^{N} \delta\big(c_n S_n^{(j)} - x\big) S_+\big(U^{(j)} - U^{(0)}\big).$$

$\tilde{G}_n^{(N)}(x, n, S_n^{(0)}, S_n(N), U(N+1))$ is in a sense an approximation of $\tilde{G}_n(x)$. Thus, it depends on the number of replicates, N, and the number of observations, n. The randomized pseudo empirical p-value function is defined as

$$\tilde{p}_n^{(N)}(x) = \frac{N\tilde{G}_n^{(N)}(x) + 1}{N+1}. \tag{5.2}$$

We can now state the Monte Carlo-based version of Theorem 5.2.

THEOREM 5.3. (*Monte Carlo test asymptotic validity*). *Let $S_n(\beta_0)$ be a test statistic for testing $H_0(\beta_0): \beta = \beta_0$ against $H_1(\beta_0): \beta \neq \beta_0$ in model (2.1) and $S_n^{(0)}$ the observed value. Suppose that, under $H_0(\beta_0)$,*

$$\mathsf{P}[c_n S_n(\beta_0) \geq x|X_n] = G_n(x|X_n) = 1 - F_n(x|X_n) \underset{n\to\infty}{\to} G(x) \; a.e.,$$

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

$$\mathsf{P}[c_n \tilde{S}_n \geq x|X_n] = \tilde{G}_n(x|X_n) = 1 - \tilde{F}_n(x|X_n) \underset{n\to\infty}{\to} G(x) \; a.e.,$$

and $(S_n^{(1)}, \ldots, S_n^{(N)})$ *be a vector of N independent replicates of* $\tilde{S}_n$, *where* $(N+1)\alpha$ *is an integer. Then, the randomized version of the Monte Carlo test with level* α *is asymptotically valid, i.e.* $\lim_{n\to\infty} \mathsf{P}[\tilde{p}_n^{(N)}(\beta_0) \leq \alpha] \leq \alpha$.

These results can be applied to the sign-based inference method. However, Theorems 5.2 and 5.3 are much more general. They do not exclusively rely on asymptotic normality—the limiting distribution may be different from a Gaussian one. Besides, the rate of convergence may differ from $\sqrt{n}$.

5.2.2. Asymptotic validity of sign-based inference. In model (2.1), suppose that conditions 5.1–5.5 hold and consider the testing problem: $H_0(\beta_0) : \beta = \beta_0$ against $H_1(\beta_0) : \beta \neq \beta_0$. Let $D_S(\beta, \hat{J}_n^{-1})$ be the test statistic as defined in (3.8). Observe $SF^{(0)} = D_S(\beta_0, \hat{J}_n^{-1})$. Draw N independent replicates of sign vector, each one having n independent components, from a $B(1, 0.5)$ distribution. Compute $(SF^{(1)}, SF^{(2)}, \ldots, SF^{(N)})$, the N pseudo replicates of $D_S(\beta_0, X'X^{-1})$ under $H_0(\beta_0)$. We call them 'pseudo' replicates because they are drawn as if observations were independent. Draw $N+1$ independent replicates $(W^{(0)}, \ldots, W^{(N)})$ from a $\mathcal{U}(0, 1)$ distribution and form the couple $(SF^{(j)}, W^{(j)})$. Compute $\tilde{p}_n^{(N)}(\beta_0)$ using (5.2). From Theorem 5.3, the confidence region $\{\beta \in \mathbb{R}^p | \tilde{p}_n^{(N)}(\beta) \geq \alpha\}$ is asymptotically conservative with level at least $1-\alpha$. $H_0(\beta_0)$ is rejected when $\tilde{p}_n^{(N)}(\beta_0) \leq \alpha$.

Contrary to usual asymptotic tests, this method does not require the existence of moments nor a density on the $\{u_t : t = 1, 2, \ldots\}$ process. Usual Wald-type inference is based on the asymptotic behaviour of estimators and, consequently, is more restrictive. More moments existence restrictions are needed; see Weiss (1991) and Fitzenberger (1997b). Besides, asymptotic variance of the LAD estimator involves the conditional density at zero of the disturbance process $\{u_t : t = 1, 2, \ldots\}$ as unknown nuisance parameter. The approximation and estimation of asymptotic covariance matrices constitute a large issue in asymptotic inference. This usually requires kernel methods. We get around those problems by adopting the finite-sample sign-based procedure.

6. SIMULATION STUDY

In this section, we study the performance of sign-based methods compared with usual asymptotic tests based on OLS or LAD estimators, with different approximations for their asymptotic covariance matrices. We consider the sign-based statistics $D_S(\beta, (X'X)^{-1})$ and $D_S(\beta, \hat{J}_n^{-1})$ when a correction is needed for linear serial dependence. We consider a set of general DGPs to illustrate different classical problems one may encounter in practice. They are presented in Table 2. First, we investigate the performance of tests, then, confidence sets. We use the following linear regression model:

$$y_t = x_t'\beta_0 + u_t, \quad t = 1, \ldots, n, \tag{6.1}$$

where $x_t = (1, x_{2,t}, x_{3,t})'$ and β_0 are 3×1 vectors. We denote the sample size n. For the first six ones, $\{u_t : t = 1, 2\ldots\}$ is i.i.d. or depends on the explanatory variables and its past values in a *multiplicative* heteroscedastic way: $u_t = h(x_t, u_{t-1}, \ldots, u_1)\varepsilon_t, t = 1, \ldots, n$. In those cases, the error term constitutes a strict conditional mediangale given X (see Assumption 2.1). Correspondingly, the levels of sign-based tests and confidence sets are perfectly controlled. Case C1 presents i.i.d. normal observations without conditional heteroscedasticity. Case C2 involves outliers in the error term. This can be seen as an example of measurement error in the observed

Table 2. Simulated models.

C1:	Normal *HOM*:	$(x_{2,t}, x_{3,t}, u_t)' \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_3),\ t = 1, \ldots, n$
C2:	Outlier:	$(x_{2,t}, x_{3,t})' \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_2)$, $u_t \overset{\text{i.i.d.}}{\sim} \begin{cases} N[0,\ 1] & \text{with } p = 0.95 \\ N[0,\ 1000^2] & \text{with } p = 0.05 \end{cases}$ $x_t, u_t,$ independent, $t = 1, \ldots, n$.
C3:	Stat. GARCH(1,1):	$(x_{2,t}, x_{3,t})' \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_2),\ u_t = \sigma_t \varepsilon_t$ with $\sigma_t^2 = 0.666 u_{t-1}^2 + 0.333 \sigma_{t-1}^2$ where $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$, $x_t, \varepsilon_t,$ independent, $t = 1, \ldots, n$.
C4:	Stoc. Volatility:	$(x_{2,t}, x_{3,t})' \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_2), u_t = \exp(w_t/2)\varepsilon_t$ with $w_t = 0.5 w_{t-1} + v_t$, where $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1),\ v_t \overset{\text{i.i.d.}}{\sim} \chi_2(3)$, $x_t, u_t,$ independent, $t = 1, \ldots, n$.
C5:	Deb. design matrix + HET. dist.:	$x_{2,t} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1),\ x_{3,t} \overset{\text{i.i.d.}}{\sim} \chi_2(1)$, $u_t = x_{3,t}\varepsilon_t,\ \varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1),\ x_t, \varepsilon_t$ independent, $t = 1, \ldots, n$.
C6:	Cauchy disturbances:	$(x_{2,t}, x_{3,t})' \sim \mathcal{N}(0, I_2)$, $u_t \overset{\text{i.i.d.}}{\sim} \mathcal{C},\ x_t, u_t,$ independent, $t = 1, \ldots, n$.
C7:	AR(1)-*HET*, $\rho_u = 0.5$, : $\rho_x = 0.5$	$x_{j,t} = \rho_x\, x_{j,t-1} + \nu_t^j, j = 1, 2$, $u_t = \min\{3, \max[0.21, \|x_{2,t}\|]\} \times \tilde{u}_t$, $\tilde{u}_t = \rho_u \tilde{u}_{t-1} + \nu_t^u$, $(\nu_t^2, \nu_t^3, \nu_t^u)' \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_3),\ t = 2, \ldots, n$ ν_1^2, ν_1^3 and ν_1^u chosen to ensure stationarity.
C8:	Exp. Var.:	$(x_{2,t}, x_{3,t}, \varepsilon_t)' \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_3), u_t = \exp(0.2t)\varepsilon_t$.

y. Cases C3 and C4 involve other non-linear dependent schemes with stationary GARCH and stochastic volatility disturbances. Case C5 combines a very unbalanced design matrix (where the LAD estimator performs poorly) with highly conditional heteroscedastic disturbances. Case C6 is an example of heavy-tailed errors (Cauchy). Next, we study the behaviour of the sign-based inference (involving a HAC correction) when inference is only asymptotically valid. Case C7 illustrates the behaviour of sign-based inference when the error term involves linear dependence at a mild level (see the discussion paper for results at other levels of linear dependence and Fitzenberger, 1997b, for a study of LAD block bootstrap performance on such DGPs). In that case, x_t and u_t are such that $E(u_t x_t) = 0$ and $E[s(u_t)x_t] = 0$ for all t. Finally, case C8 involves disturbances that are not second-order stationary (exponential variance) but for which the mediangale assumption holds. As we noted previously, sign-based inference does not require stationary assumptions in contrast with tests derived from CLT. In each case, the design matrix is simulated once. Hence, results are conditional. More simulation results on other types of DGPs can be found in the discussion paper (Coudin and Dufour, 2007).

Table 3. Linear regression under mediangale errors: empirical sizes of conditional tests for $H_0 : \beta = (1, 2, 3)'$.

$y_t = x_t\beta + u_t$,	*SIGN*		*LAD*					*OLS*		
$t = 1, \ldots, 50$.	*SF*	*SHAC*	*OS*	*DMB*	*MBB*	*BT*	*LR*	*IID*	*WH*	*BT*
			Stationary models with mediangale errors							
C1: HOM	0.052	0.050	0.086	0.050	0.089	0.047	0.068	0.060	0.096	0.113
$\rho_\epsilon = \rho_x = 0$,	*0.047**	*0.019**								
C2:	0.047	0.048	0.088	0.043	0.083	0.039	0.066	0.056	0.008	0.009
Outlier	*0.044**	*0.015**								
C3:	0.042	0.046	0.040	0.005	0.005	0.004	0.012	0.080	0.046	0.046
St. GARCH(1,1)	*0.040**	*0.013**								
C4:	0.043	0.041	0.063	0.006	0.014	0.006	0.031	0.054	0.014	0.014
Stoch. Volat.	*0.045**	*0.021**								
C5:	0.044	0.042	0.687	0.020	0.044	0.152	0.307	0.421	0.171	0.173
Deb. + Het.	*0.040**	*0.018**								
C6:	0.058	0.059	0.069	0.013	0.033	0.012	0.044	0.061	0.023	0.023
Cauchy	*0.049**	*0.021**								
			Non-stationary models with mediangale errors							
C8: Exp. Var.	0.049	0.051	0.017	0.000	0.000	0.000	0.000	0.132	0.014	0.014
			Stationary models with serial dependence							
C7: HET	0.218	0.026	0.440	0.131	0.097	0.108	0.308	0.407	0.328	0.276
$\rho_\epsilon = \rho_x = 0.5$**	–	*0.017**								

Notes: *Sizes using asymptotic critical values based on χ^2 (3).
**Automatic bandwidth parameters are restricted to be <10 to avoid invertibility problems.

6.1. Size

We first study level distortions. We consider the testing problem: $H_0(\beta_0) : \beta_0 = (1, 2, 3)'$ against $H_1 : \beta_0 \neq (1, 2, 3)'$. We compare exact and asymptotic tests based on $SF = D_S(\beta, (X'X)^{-1})$ and $SHAC = D_S(\beta, \hat{J}_n^{-1})$, where $\hat{J}_n^{-1}$ is estimated by a Bartlett kernel, with various asymptotic tests. Wald- and LR-type tests are considered. We consider Wald tests based on the OLS estimate with three different covariance estimators: the usual under homoscedasticity and independence (*IID*), White correction for heteroscedasticity (*WH*) and Bartlett kernel covariance estimator with automatic bandwidth parameter (*BT*, Andrews, 1991). Concerning the LAD estimator, we study Wald-type tests based on several covariance estimators: order statistic estimator (*OS*),[11] Bartlett kernel covariance estimator with automatic bandwidth parameter (*BT*, Powell, 1984, Buchinsky, 1995), design matrix bootstrap centring around the sample estimate (*DMB*, Buchinsky, 1998), moving block bootstrap centring around the sample estimate (*MBB*, Fitzenberger, 1997b).[12]

[11] This assumes i.i.d. residuals; an estimate of the residual density at zero is obtained from a confidence interval constructed for the $(n/2)$th residual (Buchinsky, 1998).
[12] The block size is 5.

Finally, we consider the likelihood ratio statistic (LR) assuming i.i.d. disturbances with an *OS* estimate of the error density (Koenker and Bassett, 1982). When errors are i.i.d. and X is fixed, the LM statistic for testing the joint hypothesis $H_0(\beta_0)$ turns out to be the *SF* sign-based statistic. Consequently, the three usual forms (Wald, LR and LM) of asymptotic tests are compared in our set-up.

In Table 3, we report the simulated sizes for a conditional test with nominal level $\alpha = 5\%$, given X. N replicates are used for the bootstrap and the Monte Carlo sign-based method, and $N = 2999$. All bootstrapped samples are of size $n = 50$. We simulate $M = 5000$ random samples to evaluate the sizes of these tests. For both sign-based statistics, we also report the asymptotic level whenever processes are stationary.

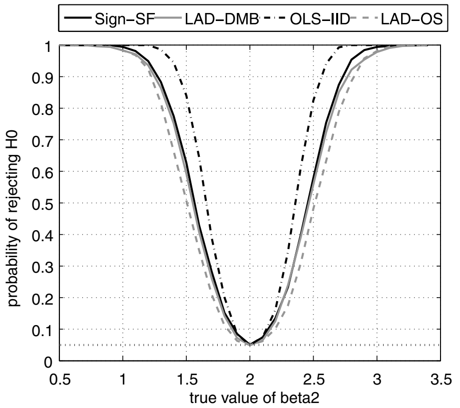
When the mediangale Assumption 2.1 holds, sizes of tests derived from sign-based finite-sample methods are exactly controlled, whereas asymptotic tests may greatly overreject or underreject the null hypothesis. This remark especially holds for cases involving strong heteroscedasticity (cases C3, C5). The asymptotic versions of sign-based tests suffer from the same underrejection than other asymptotic tests, suggesting that for small samples ($n = 50$), the distribution of the test statistic is really far from its asymptotic limit. Hence, the sign-based method that deals directly with this distribution has clearly an advantage on asymptotic methods. When the disturbance process is highly heteroscedastic (case C5), the kernel estimation of the LAD asymptotic covariance matrix is not reliable anymore.

In the last row, we illustrate behaviours when the error term involves linear serial dependence. The Monte Carlo *SHAC* sign-based test does not control exactly the level but is still asymptotically valid and yields the best results. We underscore its advantages compared with other asymptotically justified methods. Whereas the Wald and LR tests overreject the null hypothesis, the latter test seems to better control the level than its asymptotic version, avoiding underrejection. There exist important differences between using critical values from the asymptotic distribution of *SHAC* statistic and critical values derived from the distribution of the *SHAC* statistic for a fixed number of independent signs. Besides, we underscore the dramatic overrejections of asymptotic Wald tests based on HAC estimation of the asymptotic covariance matrix when the data set involves a small number of observations. These results suggest, in a sense, that when the data suffer from both a small number of observations and linear dependence, the first problem to solve is the finite-sample distortion, which is not what is usually done.

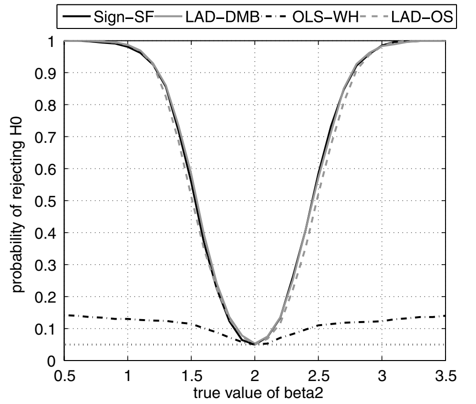
6.2. Power

Then, we illustrate the *power* of these tests. We are particularly interested in comparing the sign-based inference to kernel and bootstrap methods. We consider the simultaneous hypothesis H_0 as before. The true process is obtained by fixing β_1 and β_3 at the tested value, i.e. $\beta_1 = 1$ and $\beta_3 = 3$ and letting vary β_2. Simulated power is given by a graph with β_2 in abscissa. The power functions presented here (Figures 2 and 3) are locally adjusted for the level, which allows comparisons between methods. However, we should keep in mind that only the sign-based methods lead to exact confidence levels without adjustment. Other methods may overreject the null hypothesis and do not control the level of the test, or underreject it, and then lose power.

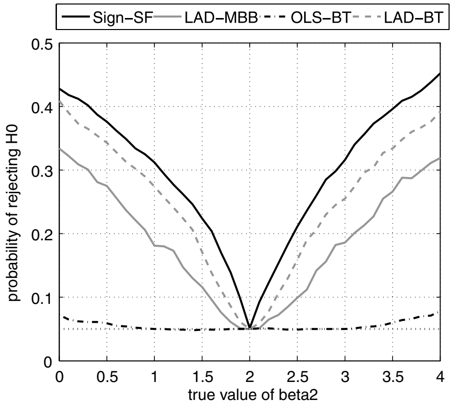
Sign-based inference has a comparable power performance with LAD methods in cases C1, C2 and a slightly lower in case C6 (Cauchy disturbances), with the advantage that the level is exactly controlled, which leads to great difference in small samples. In heteroscedastic or heterogeneous cases (C4, C5 and above all C3 and C8), sign-based inference dominates other



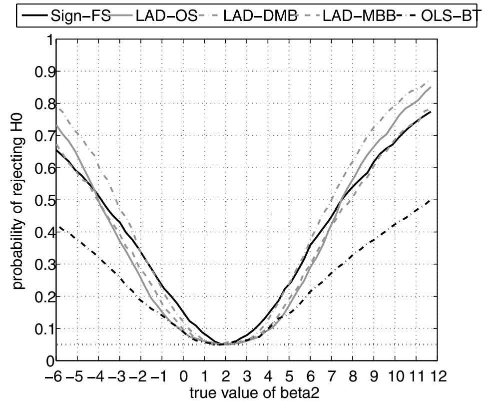
(a) C1: normal



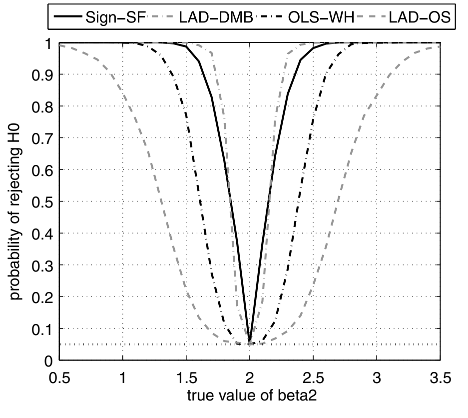
(b) C2: outliers



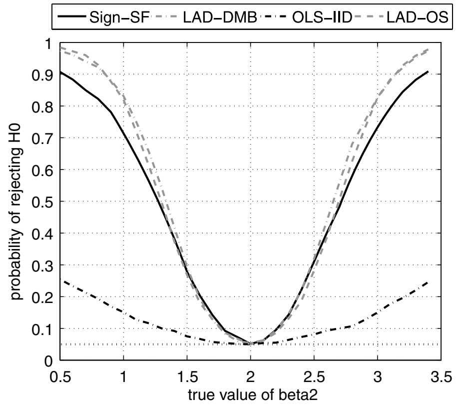
(c) C3: stationary GARCH



(d) C4: stochastic volatility

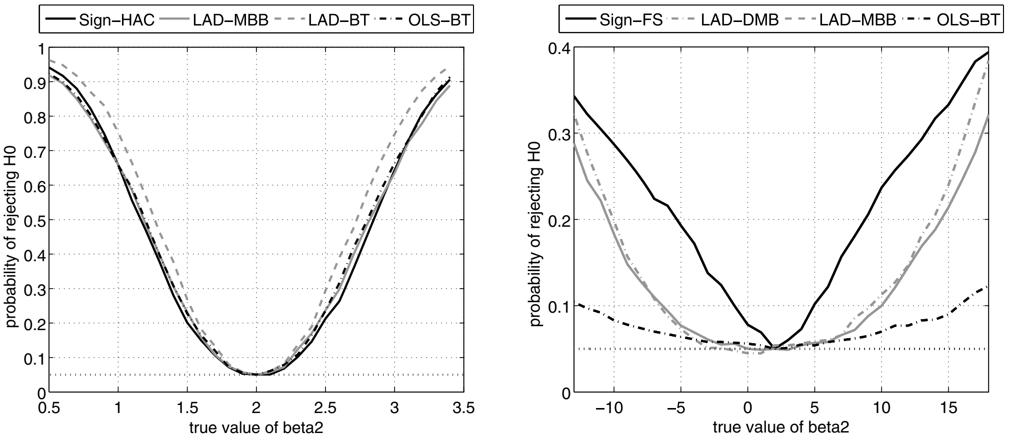


(e) C5: DEB+HET



(f) C6: Cauchy

Figure 2. Power functions (level corrected).



(a) C 7: $AR(1)$ HET, $\rho_x = \rho_u = \triangleright 5$ (b) C 12: exponential variance

Figure 3. Power functions (level corrected).

methods: levels are exactly controlled and power functions exceed others, even methods that are size-corrected with locally adjusted levels. In the presence of linear serial dependence, the Monte Carlo test based on $D_S(\beta, \hat{J}_n^{-1})$, which is still asymptotically valid, seems to lead to good power performance for a mild autocorrelation, along with a better size control (C7). Only for very high autocorrelation (close to unit root process), the sign-based inference is not adapted; see the discussion paper (Coudin and Dufour, 2007).

6.3. Confidence intervals

As the sign-based confidence regions are, by construction, of a level higher than $1 - \alpha$, whenever inference is exact, a performance indicator for confidence intervals may be their width. Thus, we wish to compare the width of confidence intervals obtained by projecting the sign-based simultaneous confidence regions to those based on t-statistics on the LAD estimator. We use $M = 1000$ simulations, and report average width of confidence intervals for each β_k and coverage probabilities in Table 4. We only consider stationary examples. Spreads of confidence intervals obtained by projection are larger than asymptotic confidence intervals. This is due to the fact that they are by construction conservative confidence intervals. However, it is not clear that valid confidence intervals without this feature can even be built.

7. ILLUSTRATIVE APPLICATION: STANDARD AND POOR'S DRIFT

We test the presence of a drift on the Standard and Poor's composite price index (*SP*), 1928–87. That process is known to involve a large amount of heteroscedasticity and have been used by Gallant et al. (1997) and Dufour and Valéry (2008) to fit a stochastic volatility model. Here, we are interested in robust testing without modelling the volatility in the disturbance process. The data set consists in a series of 16,127 daily observations of SP_t, converted to price movements, $y_t = 100[\log(SP_t) - \log(SP_{t-1})]$ and adjusted for systematic calendar effects. We consider a

Table 4. Width of confidence intervals (for stationary cases).

$y_t = x_t\beta + u_t, t = 1, \ldots, T$		Proj.-based *SF*			Proj.-based *SHAC*			LAD t-stat. with *DMB*			LAD t-stat. with *MBB*			LAD t-stat. with *BT*		
$T = 50$		β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
$(\beta_1, \beta_2, \beta_3) = (1, 2, 3)$		Models which satisfy the mediangale condition														
C1:	Average spread	1.29	1.52	1.40	1.16	1.36	1.02	0.81	0.90	0.89	0.79	0.88	0.85	0.82	0.88	0.87
$\rho_u = \rho_x = 0$	(*SD*)	(0.21)	(0.27)	(0.29)	(0.14)	(0.28)	(0.29)	(0.23)	(0.21)	(0.22)	(0.21)	(0.24)	(0.24)	(0.15)	(0.19)	(0.22)
HOM	Cov. lev.	1.0	1.0	1.0	1.0	1.0	1.0	0.97	0.97	0.97	0.95	0.96	0.95	0.97	0.96	0.96
C2:		1.26	1.37	1.05	1.15	1.24	0.91	0.92	0.94	0.98	0.88	0.98	1.04	0.88	0.88	0.88
Outlier		(0.26)	(0.31)	(0.30)	(0.25)	(0.29)	(0.30)	(0.80)	(0.79)	(1.29)	(0.67)	(1.36)	(2.73)	(0.17)	(0.20)	(0.24)
		1.0	1.0	0.98	1.0	0.99	0.96	0.98	0.98	0.98	0.97	0.97	0.97	0.97	0.98	0.97
C3:		50.4	58.5	57.3	49.5	55.9	56.1	30.6	33.4	25.9	35.0	38.3	41.5	29.3	32.6	32.3
Stat.		(101)	(118)	(122)	(100)	(115)	(117)	(64.6)	(74.6)	(61.0)	(76.7)	(82.6)	(84.0)	(70.3)	(76.9)	(78.0)
GARCH(1,1)		1.0	1.0	0.93	0.99	0.99	0.94	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
C4:		27.3	30.4	33.1	22.8	29.4	27.0	13.3	15.9	15.5	15.1	20.7	19.1	15.7	15.4	15.6
Stoc. vol.:		(14.4)	(16.7)	(18.1)	(12.2)	(17.6)	(15.8)	(6.4)	(15.9)	(15.5)	(9.6)	(28.0)	(19.3)	(7.5)	(7.8)	(7.5)
		1.0	0.98	1.0	1.0	1.0	1.0	0.99	1.0	0.99	0.98	1.0	0.99	0.99	1.0	0.99
C6:		2.20	2.75	2.59	1.88	2.33	1.95	1.25	1.47	1.47	1.21	1.41	1.42	1.39	1.52	1.53
Cauchy dist.:		(0.59)	(0.82)	(0.82)	(0.56)	(0.78)	(0.74)	(0.32)	(0.46)	(0.45)	(0.38)	(0.57)	(0.53)	(0.37)	(0.49)	(0.47)
		1.0	1.0	1.0	1.0	1.0	0.99	0.98	0.98	0.98	0.97	0.98	0.97	0.99	0.98	0.99
		Model which does not satisfy the mediangale condition														
C7:		1.25	1.46	1.56	1.23	1.64	0.99	0.68	1.12	0.96	0.79	1.23	0.96	0.94	1.11	1.01
$\rho_u = \rho_x = 0.5$		(0.31)	(0.40)	(0.40)	(0.41)	(0.51)	(0.35)	(0.17)	(0.33)	(0.24)	(0.24)	(0.42)	(0.26)	(0.33)	(0.55)	(0.36)
HET		1.0	0.99	1.0	0.98	0.97	0.94	0.93	0.88	0.98	0.95	0.89	0.98	0.97	0.83	0.97

model with a constant and a drift:

$$y_t = a + bt + u_t, \quad t = 1, \ldots, 16{,}127, \tag{7.1}$$

where we let the possibility that $\{u_t : t = 1, \ldots, 16{,}127\}$ presents a stochastic volatility or any kind of non-linear heteroscedasticity of unknown form. White and Breusch–Pagan tests for heteroscedasticity both reject homoscedasticity at 1%.[13]

We derive confidence intervals for the two parameters with the Monte Carlo sign-based method, and we compare them with the ones obtained by Wald techniques applied to LAD and OLS estimates. Then, we perform a similar experiment on two subperiods, the whole year 1929 (291 observations) and on the last 90 opened days of 1929, which roughly corresponds to the four last months of 1929 (90 observations), to investigate behaviours of the different methods in small samples. Due to the financial crisis, one may expect data to involve heavy heteroscedasticity during this period. Let us remind ourselves that the Wall Street Crash occurred between October 24 (*Black Thursday*) and October 29 (*Black Tuesday*). Hence, the second subsample corresponds to September and October with the crash period, and November and December with the early beginning of the Great Depression Heteroscedasticity tests reject homoscedasticity for both subsamples.[14]

In Table 5, we report 95% confidence intervals for a and b obtained by various methods: finite-sample sign-based method (for *SF* and *SHAC*, which involves a HAC correction); LAD and OLS with different estimates of their asymptotic covariance matrices (order statistic, bootstrap, kernel, ...). If the mediangale Assumption 2.1 holds, the sign-based confidence interval coverage probabilities are controlled. First, results on the drift are very similar between methods. The absence of a drift cannot be rejected with 5% level, but results concerning the constant differ greatly between methods and time periods. In the whole sample, the conclusions of Wald tests based on the LAD estimator differ depending on the choice of the covariance matrix estimate. Concerning the test of a positive constant, Wald tests with bootstrap or with an estimate derived if observations are i.i.d. (*OS* covariance matrix), which is totally illusory in that sample, reject, whereas the Wald test with kernel (so as sign-based tests) cannot reject the nullity of a. This may lead the practitioner in a perplex mind. Which is the correct test?

In all the considered samples, Wald tests based on OLS appear to be unreliable. Either confidence intervals are huge (see OLS results on both subperiods) or some bias is suspected (see OLS results on the whole period). Take the constant parameter, on one hand, sign-based confidence intervals and LAD confidence intervals are rather deported to the right; on the other hand, OLS confidence intervals seem to be biased towards zero. This may due to the presence of some influential observations. Moreover, the OLS estimate for the whole sample is negative. In settings with arbitrary heteroscedasticity, LS methods should be avoided.

Let us examine the third column of Table 5. The tightest confidence intervals for the constant parameter is obtained for sign-based tests based on the *SHAC* statistic, whereas LAD (and OLS) ones are larger. Note besides the gain obtained by using *SHAC* instead of *SF* in that set-up. This suggests the presence of autocorrelation in the disturbance process. In such a circumstance, finite-sample sign-based tests remain asymptotically valid, such as Wald methods. However, they are also corrected for the sample size and yield very different results. Finally, sign-based tests

[13] White: 499 (p-value = 0.000) ; BP: 2781 (p-value = 0.000).

[14] 1929: White: 24.2 (p-value = 0.000); BP: 126 (p-value = 0.000); Sept–Oct–Nov–Dec 1929: White: 11.08 (p-value = 0.004); BP: 1.76 (p-value = 0.18).

Table 5. S&P price index: 95% confidence intervals.

Constant parameter (a)	Whole sample	Subsamples	
Methods	*(16,120 obs.)*	*1929 (291 obs.)*	*1929 (90 obs.)*
Sign			
SF statistics	[−0.007, 0.105]	[−0.226, 0.522]	[−1.464, 0.491]
SHAC statistics	[−0.007, 0.106]	[−0.135, 0.443]	[−0.943, 0.362]
LAD (estimate)	(*0.062*)	(*0.163*)	(*−0.091*)
with OS cov. matrix est.	[0.033, 0.092]	[−0.144, 0.470]	[−1.015, 0.832]
with DMB cov. matrix est.	[0.007, 0.117]	[−0.139, 0.464]	[−1.004, 0.822]
with MBB cov. matrix est. (b=3)	[0.008, 0.116]	[−0.130, 0.456]	[−1.223, 1.040]
with kernel cov. matrix est. (Bn=10)	[−0.019, 0.143]	[−0.454, −0.780]	[−1.265, 1.083]
OLS	(*−0.005*)	(*0.224*)	(*−0.522*)
with i.i.d. cov. matrix est.	[−0.041, 0.031]	[−0.276, 0.724]	[−2.006, 0.962]
with DMB cov. matrix est.	[−0.054, 0.045]	[−0.142, 0.543]	[−1.335, 0.290]
with MBB cov. matrix est. (b=3)	[−0.056, 0.046]	[−0.140, 0.588]	[−1.730, 0.685]
Drift parameter (b)			
Methods	$\times 10^{-5}$	$\times 10^{-2}$	$\times 10^{-1}$
Sign			
SF statistics	[−0.676, 0.486]	[−0.342, 0.344]	[−0.240, 0.305]
SHAC statistics	[−0.699, 0.510]	[−0.260, 0.268]	[−0.204, 0.224]
LAD	(*0.184*)	(*0.000*)	(*−0.044*)
with OS cov. matrix est.	[−0.504, 0.320]	[−0.182, 0.182]	[−0.220, 0.133]
with DMB cov. matrix est.	[−0.688, 0.320]	[−0.256, 0.255]	[−0.281, 0.194]
with MBB cov. matrix est. (b=3)	[−0.681, 0.313]	[−0.236, 0.236]	[−0.316, 0.229]
with kernel cov. matrix est.	[−0.671, −0.104]	[−0.392, 0.391]	[−0.303, 0.215]
OLS	(*0.266*)	(*−0.183*)	(0.010)
with i.i.d. cov. matrix est.	[−0.119, 0.651]	[−0.480, 0.113]	[−0.273, 0.293]
with DMB cov. matrix est.	[−0.213, 0.745]	[−0.544, 0.177]	[−0.148, 0.169]
with MBB cov. matrix est. (b=3)	[−0.228, 0.761]	[−0.523, 0.156]	[−0.250, 0.270]

seem really adapted for small sample settings. Consequently, they are also particularly adapted to regional data sets, which have, by nature, fixed sample size.[15]

8. CONCLUSION

In this paper, we have proposed an entire system of inference for the β parameter of a linear median regression that relies on distribution-free sign-based statistics. We show that

[15] For an illustration on cross-regional β-convergence between the levels of per capita output in the U.S., see the discussion paper.

the procedure yields exact tests in finite samples for mediangale processes and remains asymptotically valid for more general processes, including stationary ARMA disturbances. 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 arbitrary heteroscedasticity is suspected in the data and the number of available observations is small. Finally, we have presented a practical example: we test the presence of a drift on the S&P price index, for the whole period 1928–87 and for shorter subsamples.

ACKNOWLEDGEMENTS

The authors thank Marine Carrasco, Marc Hallin, Frédéric Jouneau, Thierry Magnac, Bill McCausland, Benoit Perron, Alain Trognon, the two anonymous referees and the editor Richard Smith for useful comments and constructive discussions. Earlier versions of this paper were presented at the 2003 Meeting of the Statistical Society of Canada (Halifax), the 2005 Econometric Society World Congress (London), CREST (Paris), the 2005 Conference in honour of Jean-Jacques Laffont (Toulouse), the 2005 Workshop on 'New Trouble for Standard Regression Analysis' (Universität Regensburg, Germany) and ECARES (Brussels). This work was supported by the William Dow Chair in Political Economy (McGill University), the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Bank of Canada (Research Fellowship), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Institut de finance mathématique de Montréal (IFM2), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds de recherche sur la société et la culture (Québec).

REFERENCES

Abdelkhalek, T. and J.-M. Dufour (1998). Statistical inference for computable general equilibrium models, with application to a model of the Moroccan economy. *Review of Economics and Statistics 80*, 520–34.

Amemiya, T. (1982). Two-stage least absolute deviations estimator. *Econometrica 50*, 689–711.

Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*, 817–58.

Bahadur, R. and L. Savage (1956). The nonexistence of certain statistical procedures in nonparametric problems. *Annals of Mathematical Statistics 27*, 1115–22.

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

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

Buchinsky, M. (1995). Estimating the asymptotic covariance matrix for quantile regression models. *Journal of Econometrics 68*, 303–38.

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

Buchinsky, M. and J. Hahn (1998). An alternative estimator for the censored quantile regression model. *Econometrica 66*, 653–71.

Campbell, B. and J.-M. Dufour (1991). Over-rejections in rational expectations models: a nonparametric approach to the Mankiw–Shapiro problem. *Economics Letters 35*, 285–90.

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

Campbell, B. and J.-M. Dufour (1997). Exact nonparametric tests of orthogonality and random walk in the presence of a drift parameter. *International Economic Review 38*, 151–73.

Chernozhukov, V., C. Hansen and M. Jansson (2008). Finite sample inference for quantile regression models. Forthcoming in *Journal of Econometrics*.

Chow, Y. S. and H. Teicher (1988). *Probability Theory. Independence, Interchangeability, Martingales* (2nd ed.). New York: Springer-Verlag.

Coudin, E. and J.-M. Dufour (2006). Generalized confidence distributions and robust sign-based estimators in median regressions under heteroskedasticity and nonlinear dependence of unknown form. Working paper, CIRANO-CIREQ, McGill University.

Coudin, E. and J.-M. Dufour (2007). Finite and large sample distribution-free inference in linear median regressions under heteroskedasticity and nonlinear dependence of unknown form. Working paper, CIREQ, McGill University and Document de Travail No. 2007-38, CREST-INSEE.

David, H. A. (1981). *Order Statistics* (2nd ed.). New York: John Wiley.

De Angelis, D., P. Hall and G. A. Young (1993). Analytical and bootstrap approximations to estimator distributions in L_1 regression. *Journal of the American Statistical Association 88*, 1310–16.

Dielman, T. and R. Pfaffenberger (1988a). Bootstrapping in least absolute value regression: an application to hypothesis testing. *Communications in Statistics—Simulation and Computation 17*, 843–56.

Dielman, T. and R. Pfaffenberger (1988b). Least absolute value regression: necessary sample sizes to use normal theory inference procedures. *Decision Sciences 19*, 734–43.

Dodge, Y. (Ed.) (1997). *L_1-Statistical Procedures and Related Topics*, Lecture Notes—Monograph Series, Volume 31. Hayward, CA: Institute of Mathematical Statistics.

Dubrovin, B., A. Fomenko and S. Novikov (1984). *Modern Geometry—Methods and Applications*. New York: Springer-Verlag.

Dufour, J.-M. (1981). Rank tests for serial dependence. *Journal of Time Series Analysis 2*, 117–28.

Dufour, J.-M. (1990). Exact tests and confidence sets in linear regressions with autocorrelated errors. *Econometrica 58*, 475–94.

Dufour, J.-M. (1997). Some impossibility theorems in econometrics, with applications to structural and dynamic models. *Econometrica 65*, 1365–89.

Dufour, J.-M. (2006). Monte Carlo tests with nuisance parameters: a general approach to finite-sample inference and nonstandard asymptotics in econometrics. *Journal of Econometrics 133*, 443–77.

Dufour, J.-M., M. Hallin and I. Mizera (1998). Generalized runs tests for heteroskedastic time series. *Journal of Nonparametric Statistics 9*, 39–86.

Dufour, J.-M. and J. Jasiak (2001). Finite sample limited information inference methods for structural equations and models with generated regressors. *International Economic Review 42*, 815–43.

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

Dufour, J.-M. and M. Taamouti (2005). Projection-based statistical inference in linear structural models with possibly weak instruments. *Econometrica 73*, 1351–65.

Dufour, J.-M. and P. Valéry (2008). Exact and asymptotic tests for possibly non-regular hypotheses on stochastic volatility models. Forthcoming in *Journal of Econometrics*.

Dwass, M. (1957). Modified randomization tests for nonparametric hypotheses. *Annals of Mathematical Statistics 28*, 181–87.

Fitzenberger, B. (1997a). A guide to censored quantile regressions. In G. S. Maddala and C. R. Rao (Eds.), *Handbook of Statistics*, *Volume 15*, 405–37. Amsterdam: North-Holland.

Fitzenberger, B. (1997b). The moving blocks bootstrap and robust inference for linear least squares and quantile regressions. *Journal of Econometrics 82*, 235–87.

Gallant, A. R., D. Hsieh and G. Tauchen (1997). Estimation of stochastic volatility models with diagnostics. *Journal of Econometrics 81*, 159–92.

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

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

Gray, A. (1998). *Modern Differential Geometry of Curves and Surfaces with Mathematica* (2nd ed.). Boca Raton, FL: CRC Press.

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

Hallin, M. and M. L. Puri (1991). Time series analysis via rank-order theory: signed-rank tests for ARMA models. *Journal of Multivariate Analysis 39*, 175–237.

Hallin, M. and M. L. Puri (1992). Rank tests for time series analysis: a survey. In D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt and M. S. Taqqu (Eds.), *New Directions in Time Series Analysis (Part I)*, The IMA Volumes in Mathematics and its Applications, Volume 45, 111–53. New York: Springer-Verlag.

Hallin, M., C. Vermandele and B. Werker (2006). Linear and nonserial sign-and-rank statistics: asymptotic representation and asymptotic normality. *The Annals of Statistics 34*, 254–89.

Hallin, M., C. Vermandele and B. Werker (2008). Semiparametrically efficient inference based on signs and ranks for median-restricted models. *Journal of the Royal Statistical Society*, Series B *70*, 389–412.

Hallin, M. and B. Werker (2003). Semiparametric efficiency, distribution-freeness, and invariance. *Bernoulli 9*, 137–65.

Hong, H. and E. Tamer (2003). Inference in censored models with endogenous regressors. *Econometrica 71*, 905–32.

Horowitz, J. L. (1998). Bootstrap methods for median regression models. *Econometrica 66*, 1327–51.

Jung, S. (1996). Quasi-likelihood for median regression models. *Journal of the American Statistical Association 91*, 251–57.

Kim, T. and H. White (2002). Estimation, inference, and specification testing for possibly misspecified quantile regression. Discussion paper 2002-09, Department of Economics, UC San Diego.

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

Koenker, R. and G. Bassett, Jr. (1982). Tests of linear hypotheses and L_1 estimation. *Econometrica 50*, 1577–84.

Komunjer, I. (2005). Quasi-maximum likelihood estimation for conditional quantiles. *Journal of Econometrics 128*, 137–64.

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

Linton, O. and Y.-J. Whang (2007). The quantilogram: with an application to evaluating directional predictability. *Journal of Econometrics 141*, 250–82.

Mankiw, G. and M. Shapiro (1986). Do we reject too often? Small samples properties of tests of rational expectations models. *Economics Letters 20*, 139–45.

Newey, W. and K. West (1987). A simple positive semi-definite, heteroskadasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–8.

Parzen, E. (1957). On consistent estimates of the spectrum of a stationary times series. *Annals of Mathematical Statistics 28*, 329–48.
Powell, J. L. (1983). The asymptotic normality of two-stage least absolute deviations estimators. *Econometrica 51*, 1569–76.
Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics 25*, 303–25.
Powell, J. L. (1986). Censored regression quantiles. *Journal of Econometrics 32*, 143–55.
Powell, J. L. (1994). Estimation of semiparametric models. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, *Volume 4*, 2443–521. Amsterdam: North-Holland.
Pratt, J. W. and J. D. Gibbons (1981). *Concepts of Nonparametric Theory*. New York: Springer-Verlag.
Press, W. H., S. A. Teukolsky, W. T. Vetterling and B. P. Flannery (1996). *Numerical Recipes in Fortran 90* (2nd ed.). Cambridge: Cambridge University Press.
Scheffé, H. and J. W. Tukey (1945). Non-parametric estimation, I: validation of order statistics. *Annals of Mathematical Statistics 116*, 187–92.
Thompson, W. R. (1936). On confidence ranges for the median and other expectation distributions for populations of unknown distribution form. *Annals of Mathematical Statistics 7*, 122–8.
Weiss, A. (1991). Estimating nonlinear dynamic models using least absolute error estimation. *Econometric Theory 7*, 46–68.
White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica 48*, 817–38.
White, H. (2001). *Asymptotic Theory For Econometricians* (revised edition). New York: Academic Press.
Wright, J. H. (2000). Alternative variance-ratio tests using ranks and signs. *Journal of Business and Economic Statistics 18*, 1–9.
Zhao, Q. (2001). Asymptotically efficient median regression in presence of heteroskedasticity of unknown form. *Econometric Theory 17*, 765–84.

APPENDIX A: PROOFS

Proof of Proposition 2.1: We use the fact that, as $\{X_t : t = 1, 2, \dots\}$ is strongly exogenous, $\{u_t : t = 1, 2, \dots\}$ does not Granger cause $\{X_t : t = 1, 2, \dots\}$. It follows directly that $l(s_t \,|u_{t-1}, \dots, u_1, x_t, \dots, x_1) = l(s_t|u_{t-1}, \dots, u_1, x_n, \dots, x_1)$, where l stands for the density of $s_t = s(u_t)$. □

Proof of Proposition 3.1: Consider the vector $[s(u_1),\ s(u_2), \dots, s(u_n)]' \equiv (s_1,\ s_2, \dots, s_n)'$. From Assumption 2.1, we derive the two following equalities:

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

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

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

Hence, conditional on X, $s_1, s_2, \dots, s_n \overset{\text{i.i.d.}}{\sim} \mathcal{B}(1/2)$. □

Proof of Proposition 3.2: Consider model (2.1) with $\{u_t : t = 1, 2, \ldots\}$, satisfying a weak mediangale conditional on X. Let us show that $\tilde{s}(u_1), \tilde{s}(u_2), \ldots, \tilde{s}(u_n)$ can have the same role in Proposition 3.1 as $s(u_1), s(u_2), \ldots, s(u_n)$ under Assumption 2.1. The randomized signs are defined by $\tilde{s}(u_t, V_t) = s(u_t) + [1 - s(u_t)^2]s(V_t - 0.5)$, hence

$$\mathsf{P}[\tilde{s}(u_t, V_t) = 1 | u_{t-1}, \ldots, u_1, X] = \mathsf{P}\left[s(u_t) + \left[1 - s(u_t)^2\right] s(V_t - 0.5) = 1 | u_{t-1}, \ldots, u_1, X\right].$$

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

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

The weak conditional mediangale assumption given X entails

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

where $p_t = \mathsf{P}[u_t = 0 | u_{t-1}, \ldots, u_1, X]$. Substituting (A.2) into (A.1) yields

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

In a similar way,

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

The rest is similar to the proof of Proposition 3.1. □

Proof of Proposition 3.3: Let us consider first, the case of a single explanatory variable case ($p = 1$), which contains the basic idea for the proof. The case with $p > 1$ is just an adaptation of the same ideas to multidimensional notions. Under model (2.1) with the mediangale Assumption 2.1, the locally optimal sign-based test (conditional on X) of $H_0(\beta) : \beta = 0$ against $H_1(\beta) : \beta \neq 0$ is well defined. Among tests with level α, the power function of the locally optimal sign-based test has the highest slope around zero. The power function of a sign-based test conditional on X can be written $\mathsf{P}_\beta[s(y) \in W_\alpha | X]$, where W_α is the critical region with level α. Hence, we should include in W_α the sign vectors for which $\frac{d}{d\beta}\mathsf{P}_\beta[S(y) = s | X]_{\beta=0,}$ is as large as possible. An easy way to determine that derivative is to identify the terms of a Taylor expansion around zero. Under Assumption 2.1, we have

$$\mathsf{P}_\beta[S(y) = s | X] = \prod_{i=1}^{n} [\mathsf{P}_\beta(y_i > 0 | X)]^{(1+s_i)/2} [\mathsf{P}_\beta(y_i < 0 | X)]^{(1-s_i)/2} \quad \text{(A.5)}$$

$$= \prod_{i=1}^{n} [1 - F_i(-x_i\beta | X)]^{(1+s_i)/2} [F_i(-x_i\beta | X)]^{(1-s_i)/2}. \quad \text{(A.6)}$$

Assuming that continuous densities at zero exist, a Taylor expansion at order one entails

$$\mathsf{P}_\beta[S(y) = s | X] = \frac{1}{2^n} \prod_{i=1}^{n} [1 + 2 f_i(0 | X) x_i s_i \beta + o(\beta)] \quad \text{(A.7)}$$

$$= \frac{1}{2^n} \left[1 + 2 \sum_{i=1}^{n} f_i(0 | X) x_i s_i \beta + o(\beta)\right]. \quad \text{(A.8)}$$

All other terms of the product decomposition are negligible or equivalent to $o(\beta)$. That allows us to identify the derivative at $\beta = 0$:

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

Therefore, the required test has the form

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

or equivalently, $W_\alpha = \{s|s(y)'\tilde{X}\tilde{X}'s(y) > c'_\alpha\}$, where c_α and c'_α are defined by the significance level. When the disturbances have a common conditional density at zero, $f(0|X)$, we find the results of Boldin et al. (1997). The locally optimal sign-based test is given by $W_\alpha = \{s|s(y)' XX's(y) > c'_\alpha\}$. 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 zero for a multidimensional parameter. Boldin et al. (1997) propose to restrict to the class of locally unbiased tests with given level α and to consider the maximal mean curvature. Thus, a locally unbiased sign-based test satisfies, $\left.\frac{d\mathsf{P}_\beta(W_\alpha)}{d\beta}\right|_{\beta=0} = 0$, and, as $f'_i(0) = 0, \forall\, i$, the behaviour of the power function around zero is characterized by the quadratic term of its Taylor expansion

$$\beta'\frac{1}{2}\left(\frac{d^2\mathsf{P}_\beta(W_\alpha)}{d\beta^2}\right)\beta = \frac{1}{2^{n-2}}\sum_{1\le i\ne}\sum_{j\le n}[f_i(0|X)s_i\beta' x_i][f_j(0|X)s_j x'_j\beta]. \tag{A.11}$$

The locally most powerful sign-based test in the sense of the mean curvature maximizes the mean curvature, which is, by definition, proportional to the trace of $\left.\frac{d^2\mathsf{P}_\beta(W_\alpha)}{d\beta^2}\right|_{\beta=0}$; see Boldin et al. (1997, p. 41), Dubrovin et al. (1984, ch. 2, pp. 76–86) or Gray (1998, ch. 21, pp. 373–80). Taking the trace in expression (A.11), we find (after some computations) that

$$\mathrm{tr}\left(\left.\frac{d^2\mathsf{P}_\beta(W_\alpha)}{d\beta^2}\right|_{\beta=0}\right) = \sum_{1\le i\ne}\sum_{j\le n} f_i(0|X)f_j(0|X)s_i s_j \sum_{k=1}^{p} x_{ik}x_{jk}\,. \tag{A.12}$$

By adding the independent of s quantity $\sum_{i=1}^{n}\sum_{k=1}^{p} x_{ik}^2$ to (A.12), we find

$$\sum_{k=1}^{p}\left(\sum_{i=1}^{n} x_{ik} f_i(0|X)s_i\right)^2 = s'(y)\tilde{X}\tilde{X}'s(y)\,. \tag{A.13}$$

Hence, the locally optimal sign-biased test, in the sense developed by Boldin et al. (1997) for heteroscedastic signs, is $W_\alpha = \{s : s'(y)\tilde{X}\tilde{X}'s(y) > c'_\alpha\}$. Another quadratic test statistic convenient for large-sample evaluation is obtained by standardizing by $\tilde{X}'\tilde{X}$: $W_\alpha = \{s : s'(y)\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'s(y) > c'_\alpha\}$. □

Proof of Theorem 5.1: This proof follows the usual steps of an asymptotic normality result for mixing processes (see White, 2001). Consider model (2.1). In the following, s_t stands for $s(u_t)$. Under Assumption 5.4, $V_n^{-1/2}$ exists for any n. Set $\mathcal{Z}_{nt} = \lambda' V_n^{-1/2}x'_t s(u_t)$, for some $\lambda \in \mathbb{R}^p$ such that $\lambda'\lambda = 1$. The mixing property 5.1 of (x'_t, u_t) gets transmitted to $\mathcal{Z}_{nt}$; see White (2001, Theorem 3.49). Hence, $\lambda' V_n^{-1/2}s(u_t) \otimes x_t$ is α-mixing of size $-r/(r-2), r > 2$. Assumptions 5.2 and 5.3 imply

$$\mathsf{E}[\lambda' V_n^{-1/2}x'_t s(u_t)] = 0,\ t = 1, \ldots, n,\ \forall n \in \mathbb{N}\,. \tag{A.14}$$

$$\mathsf{E}\left|\lambda' V_n^{-1/2}x'_t s(u_t)\right|^r < \Delta < \infty,\ t = 1, \ldots, n,\ \forall n \in \mathbb{N}\,. \tag{A.15}$$

Note also that

$$\operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathcal{Z}_{nt}\right)=\operatorname{Var}\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\lambda' V_n^{-1/2}s(u_t)\otimes x_t\right]=\lambda' V_n^{-1/2}V_nV_n^{-1/2}\lambda=1\,. \tag{A.16}$$

The mixing property of $\mathcal{Z}_{nt}$ and equations (A.14)–(A.16) allow one to apply a central limit theorem (see White, 2001, Theorem 5.20) that yields

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\lambda' V_n^{-1/2}s(u_t)\otimes x_t\to\mathcal{N}(0,1)\,. \tag{A.17}$$

Since λ is arbitrary with $\lambda'\lambda=1$, the Cramér–Wold device entails

$$V_n^{-1/2}n^{-1/2}\sum_{t=1}^{n}s(u_t)\otimes x_t\to\mathcal{N}(0,I_p)\,. \tag{A.18}$$

Finally, Assumption 5.5 states that Ω_n is a consistent estimate of V_n^{-1}. Hence,

$$n^{-1/2}\Omega_n^{1/2}\sum_{t=1}^{n}s(u_t)\otimes x_t\to\mathcal{N}(0,I_p), \tag{A.19}$$

and $n^{-1}s'\,(y-X\beta_0)X\Omega_n\,X's(y-X\beta_0)\to\chi^2\,(p)$. □

Proof of Corollary 5.1: Let $\mathcal{F}_t=\sigma(y_0,\ldots,y_t,x_0',\ldots,x_t')$. When the mediangale Assumption 2.1 holds, $\{s(u_t)\otimes x_t,\mathcal{F}_t:t=1,\ldots,n\}$ belong to a martingale difference with respect to $\mathcal{F}_t$. Hence, $V_n=\operatorname{Var}[\frac{1}{\sqrt{n}}s\otimes X]=\frac{1}{n}\sum_{t=1}^{n}E(x_ts_ts_t'x_t')=\frac{1}{n}\sum_{t=1}^{n}E(x_tx_t')=\frac{1}{n}E(X'X)$, and $X'X/n$ is a consistent estimate of $E(X'X/n)$. Theorem 5.1 yields $SF(\beta_0)\to\chi_2(p)$. □

In order to prove Theorem 5.2, we will use the following lemma on the uniform convergence of distribution functions (see Chow and Teicher, 1988, sec. 8.2, p. 265).

LEMMA 8.1. *Let $(F_n)_{n\in\mathbb{N}}$ and F be right continuous distribution functions. Suppose that* $F_n(x)\underset{n\to\infty}{\to}F(x)$, $\forall x\in\mathbb{R}$. *Then,* $\sup_{-\infty<x<+\infty}|F_n(x)-F(x)|\underset{n\to\infty}{\to}0$.

Proof of Theorem 5.2: $G(-\infty)=\tilde{G}_n(-\infty)=0$, $G(+\infty)=\tilde{G}_n(+\infty)=1$, and $\tilde{G}_n(x|X_n(\omega))\to G(x)$ *a.e.* By Lemma 8.1, $(\tilde{G}_n)_{n\mathbb{N}}$ converges uniformly to G. The same holds for G_n. Moreover, $\tilde{G}_n$ can be rewritten as

$$\begin{aligned}\tilde{G}_n(c_nS_n(\beta_0)|X_n)&=[\tilde{G}_n(c_nS_n(\beta_0)|X_n(\omega))-G(c_nS_n(\beta_0))]\\&\quad+[G(c_nS_n(\beta_0))-G_n(c_nS_n(\beta_0)|X_n(\omega))]+G_n(c_nS_n(\beta_0)|X_n),\end{aligned}$$

hence

$$G_n(c_nS_n(\beta_0)|X_n)=\tilde{G}_n(c_nS_n(\beta_0)|X_n)+o_p(1). \tag{A.20}$$

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

$$\mathsf{P}[G_n(c_nS_n(\beta_0))\le\alpha]=\alpha. \tag{A.21}$$

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

$$\mathsf{P}[G_n(c_nS_n(\beta_0))\le\alpha_1]\le\mathsf{P}[G_n(c_nS_n(\beta_0))\le\alpha]\le\mathsf{P}[G_n(c_nS_n(\beta_0))\le\alpha_2].$$

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

$$\begin{aligned}\mathsf{P}[G_n(c_nS_n(\beta_0)) \le \alpha] &= \mathsf{P}[\{G_n(c_nS_n(\beta_0)) \le \alpha_1\} \cup \{\alpha_1 < G_n(c_nS_n(\beta_0)) \le \alpha\}] \\ &= \mathsf{P}[G_n(c_nS_n(\beta_0)) \le \alpha_1] + 0,\end{aligned}$$

as $\{\alpha_1 < G_n(c_nS_n(\beta_0)) \le \alpha\}$ is a zero-probability event. Applying (A.21) to α_1,

$$\mathsf{P}[G_n(c_nS_n(\beta_0)) \le \alpha] = \mathsf{P}[G_n(c_nS_n(\beta_0)) \le \alpha_1] = \alpha_1 \le \alpha. \tag{A.22}$$

Hence, for $\alpha \in (0, 1)$, we have $\mathsf{P}[G_n(c_nS_n(\beta_0)) \le \alpha] \le \alpha$. The latter combined with equation (A.20) allows us to conclude

$$\mathsf{P}[\tilde{G}_n(c_nS_n(\beta_0)) \le \alpha] = \mathsf{P}[G_n(c_nS_n(\beta_0)) \le \alpha] + o_p(1) \le \alpha + o_p(1). \qquad \square$$

Proof of Theorem 5.3: Let $S_n^{(0)}$ be the observed statistic and $S_n(N) = (S_n^{(1)}, \ldots, S_n^{(N)})$, a vector of N independent replicates drawn from $\tilde{F}_n(x)$. Usually, validity of Monte Carlo testing is based on the fact the vector $(c_nS_n^{(0)}, \ldots, c_nS_n^{(N)})$ is exchangeable. Indeed, in that case, the distribution of ranks is fully specified and yields the validity of empirical p-value (see Dufour, 2006). In our case, it is clear that $(c_nS_n^{(0)}, \ldots, c_nS_n^{(N)})$ is not exchangeable, so that Monte Carlo validity cannot be directly applied. Nevertheless, asymptotic exchangeability still holds, which will enable us to conclude. To obtain that the vector $(c_nS_n^{(0)}, \ldots, c_nS_n^{(N)})$ is asymptotically exchangeable, we show that for any permutation $\pi : [1, N] \to [1, N]$,

$$\lim_{n\to\infty} \mathsf{P}\left[S_n^{(0)} \ge t_0, \ldots, S_n^{(N)} \ge t_N\right] - \mathsf{P}\left[S_n^{\pi(0)} \ge t_0, \ldots, S_n^{\pi(N)} \ge t_N\right] = 0.$$

First, let us rewrite

$$\mathsf{P}\left[S_n^{(0)} \ge t_0, \ldots, S_n^{(N)} \ge t_N\right] = \mathsf{E}_{X_n}\left\{\mathsf{P}\left[S_n^{(0)} \ge t_0, \ldots, S_n^{(N)} \ge t_N, X_n = x_n\right]\right\}.$$

The conditional independence of the sign vectors (replicated and observed) entails:

$$\begin{aligned}\mathsf{P}\left[S_n^{(0)} \ge t_0, \ldots, S_n^{(N)} \ge t_N, X_n = x_n\right] &= \mathsf{P}[X_n = x_n] \prod_{i=0}^{N} \mathsf{P}\left[S_n^{(i)} \ge t_i | X_n = x_n\right] \\ &= G_n(t_0|X_n = x_n) \prod_{i=1}^{N} \tilde{G}_n(t_i|X_n = x_n).\end{aligned}$$

As each survival function converges with probability one to $G(x)$, we finally obtain

$$\mathsf{P}\left[S_n^{(0)} \ge t_0, S_n^{(1)} \ge t_1, \ldots, S_n^{(N)} \ge t_N, X_n = x_n\right] \to \prod_{i=0}^{N} G(t_i) \text{ with probability one.}$$

Moreover, it is straightforward to see that for $\pi : [1, N] \to [1, N]$, we have, as $n \to \infty$,

$$\mathsf{P}\left[S_n^{(0)} \ge t_{\pi(0)},\ S_n^{\pi(1)} \ge t_1, \ldots, S_n^{\pi(N)} \ge t_N,\ X_n = x_n\right] \to \prod_{i=0}^{N} G(t_i) \text{ with probability one.}$$

$G(t)$ is not a function of the realization $X(\omega)$, so that

$$\lim_{n\to\infty} \mathsf{P}\left[S_n^{(0)} \ge t_0, \ldots, S_n^{(N)} \ge t_N\right] - \mathsf{P}\left[S_n^{\pi(0)} \ge t_0, \ldots, S_n^{\pi(N)} \ge t_N\right] = 0.$$

Hence, we can apply an asymptotic version of proposition 2.2.2 in Dufour (2006) 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 (2006). We just have to replace the exact distributions of randomized ranks, the empirical survival functions and the empirical p-values by their asymptotic counterparts, and this is sufficient to conclude. Suppose that N, the number of replicates is such that $\alpha(N + 1)$ is an integer. Then, $\lim_{n\to\infty} \tilde{p}_n^N(c_nS_n^{(0)}) \le \alpha$. $\square$