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

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

First version: February 2003
Revised: April 2004, February 2005, November 2006,
March 2007, March 2008
This version: April 1, 2008
Compiled: April 1, 2008, 1:51am

* The authors thank Marine Carrasco, Frédéric Jouneau, Thierry Magnac, Bill McCausland, Benoit Perron, and Alain Trognon for useful comments and constructive discussions. Earlier versions of this paper were presented at 2003 Meeting of 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, the Fonds de recherche sur la société et la culture (Québec), and the Fonds de recherche sur la nature et les technologies (Québec), and a Killam Fellowship (Canada Council for the Arts).

† Institut national de la statistique et des études économiques (INSEE), and Centre de recherche en économie et statistique (CREST). Mailing address: Laboratoire de microéconometrie, CREST, Timbre J390, 15 Bd G. Péri, 92245 Malakoff Cedex, France. E-mail: elise.coudin@insee.fr.

‡ William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 519, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 8879; FAX: (1) 514 398 4938; e-mail: jean-marie.dufour@mcgill.ca . Web page: http://www.jeanmariedufour.com

ABSTRACT

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 setup studied allows for nonnormality, both discrete and continuous distributions, heteroskedasticity and nonlinear serial dependence of unknown forms. 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 nonlinear 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 even 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), for example by allowing the absence of finite moments as well as discrete distributions to which usual density estimation methods do not apply. Simulation results illustrate the performance of the procedures. Finally, the proposed methods are applied to two empirical examples: a test of the drift in the Standard and Poor's composite price index series (allowing for conditional heteroskedasticity of unknown form), and a test of β-convergence between levels of per capita output across U.S. States.

Key words: sign-based methods; median regression; simultaneous inference; Monte Carlo tests; bootstrap; projection methods; quantile regressions; non-normality; heteroskedasticity; serial dependence; GARCH; stochastic volatility; sign test.
Journal of Economic Literature classification: C12, C14, C15.

RÉSUMÉ

Dans cet article, nous construisons des tests et des régions de confiance pour les paramètres d'une régression linéaire sur la médiane, qui sont valides à distance finie sans imposer d'hypothèse paramétrique sur la distribution des erreurs. Les erreurs peuvent être non gaussiennes, hétéroscédastique ou bien, présenter une dépendance sérielle de forme arbitraire. Habituellement, l'analyse de ces modèles semi-paramétriques s'appuie sur des approximations asymptotiques normales, lequelles peuvent être trompeuses en échantillon fini. Nous introduisons une propriété analogue à la différence de martingale pour la médiane, la «médiangale» et remarquons que les signes d'une suite de «médiangale» sont indépendants entre eux et suivent une distribution connue et simulable. Nous utilisons la transformation par les signes et proposons des statistiques pivotales qui, en plus d'être robustes, permettent de construire une approche d'inférence simultanée valide quelle que soit la taille de l'échantillon. Nous utilisons la méthode des tests de Monte Carlo, puis déduisons par projection des tests et des régions de confiance pour n'importe quelle transformation du paramètre. Nous fournissons aussi une théorie asymptotique sous des hypothèses plus faibles. Les études par simulation illustrent la performance de la méthode proposée lorsque les données sont très hétérogènes. Enfin, nous présentons deux exemples d'application.

Mots clés : méthodes de signes ; régression sur la médiane ; échantillons finis; non normalité ; hétéroscédasticité ; dépendance sérielle ; GARCH ; volatilité stochastique ; tests de signes ; inférence simultanée ; tests de Monte Carlo ; bootstrap ; méthodes de projection ; régressions quantiles.
Classification JEL : C12, C14, C15.

Contents

List of Definitions, Propositions and Theorems

List of Tables

1. Introduction

The Laplace-Boscovich median regression has attracted renewed interest in recent years, especially because it is considerably more robust to non-normality and outliers than least squares [see Dodge (1997)]. It has been adapted to models involving heteroskedasticity and autocorrelation [Zhao (2001), Weiss (1990)], endogeneity [Amemiya (1982), Powell (1983), Hong and Tamer (2003)], nonlinear functional forms [Weiss (1991)] and has been extended to quantile regressions [Koenker and Bassett (1978)]. Theoretical advances on the behavior of the associated estimators have completed this process [Powell (1994), Chen, Linton and Van Keilegom (2003)]. In empirical studies, partly thanks to the generalization to quantile regressions, new fields of potential applications have emerged.[1] The recent and fast development of computer technology clearly stimulates interest for these robust, but formerly cumbersome, methods.

Linear median regression assumes a linear relation between the dependent variable y and the explanatory variables x. Only a null median assumption is imposed on the disturbance process. Such a condition of identification "by the median" can be motivated by fundamental results on nonparametric inference. Since Bahadur and Savage (1956), it is known that without strong distributional assumptions (such as normality), it is impossible to obtain reasonable tests on the mean of independent identically distributed (*i.i.d.*) observations, for any sample size. In general, moments are not empirically meaningful without further distributional assumptions. This form of nonidentification can be eliminated by choosing alternative measures of central tendency, such as the median, because nonparametric hypotheses on the median can be tested through signs tests. This suggests that median identification is more appropriate in nonparametric setups than its mean counterpart.

Median regression (and related quantile regressions) provides an attractive bridge between parametric and nonparametric 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. This holds for heavy-tailed distributions or distributions with a probability mass at zero. 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 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 for a linear median regression model. The family of statistics con-

[1] The reader is referred to Buchinsky (1994) for an interpretation in terms of inequality and mobility topics in the U.S. labor market, Engle and Manganelli (2000) for an application in Value at Risk issues in finance. For reviews of this literature, see Buchinsky (1998), Koenker and Hallock (2001) and Koenker (2005).

sidered include optimal sign tests. 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 heteroskedasticity and nonlinear dependence of unknown forms, as well as for *discrete* distributions. This is done in particular by combining Monte Carlo tests adapted to discrete distributions (using a tie-breaking procedure) with projection techniques (to allow inference on general parameter transformations). We also show that the tests proposed include locally optimal tests. Second, under even 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). For example, moment non-existence is allowed as well as discrete distributions (to which density estimators required by LAD tests do not apply).

A basic motivation for the sign-based methods considered 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 heteroskedasticity 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 valid inference for any sample size. Second, all other methods, including the usual heteroskedasticity and autocorrelation corrected (HAC) methods developed by White (1980), Newey and West (1987), Andrews (1991) and others, which are not based on signs, are not proved to be valid for any sample size. Although this provides a compelling argument for using sign-based procedures, the latter have barely been exploited in econometrics. Our point is to stress their robustness and to generalize their use to median regressions.

To our knowledge, sign-based methods have not received much attention 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, Simonova and Tyurin (1997). These authors also develop sign-based inference and estimation for linear models, both exact and asymptotic with *i.i.d.* errors. We consider sign-based statistics related to locally optimal sign tests, which are simple quadratic forms and can easily be used for estimation as well. However, we demonstrate this distribution-free property to allow for a wide array of nonlinear dependence schemes. An important feature of these 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 [Dufour (1981), Campbell and Dufour (1991, 1995, 1997)]. We combine them with projection techniques and Monte Carlo tests to derive exact confidence sets.

The pivotality of the sign-based statistics validates the use of Monte Carlo tests, a technique proposed by Dwass (1957) and Barnard (1963). This method, once adapted to discrete distributions by a tie-breaking procedure [Dufour (2006)], yields exact simultaneous confidence regions for β. Then, conservative confidence intervals (CIs) for each component of the parameter (or any real function of the parameter) can be obtained by projection [Dufour (1990), Dufour and Kiviet (1998), Dufour and Jasiak (2001), Dufour and Taamouti (2005)]. In particular, confidence interval (or set) boundaries may be calculated using global optimization methods such as simulated annealing [see Goffe, Ferrier and Rogers (1994)].

Sign-based inference methods constitute an alternative to inference derived from the asymptotic distribution of LAD estimators. The LAD estimator (such as related quantile estimators) is consistent and asymptotically normal in case of heteroskedasticity [Powell (1984) and Zhao (2001) for efficient weighted LAD estimator], or temporal dependence [Weiss (1991)]. Fitzenberger (1997*b*) extended the scheme of potential temporal dependence including stationary ARMA disturbance processes. Horowitz (1998) proposed a smoothed version of the LAD estimator. At the same time, 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) advocated the use of design matrix bootstrap for independent observations. In dependent cases, Fitzenberger (1997*b*) proposed a moving block bootstrap. Finally, Hahn (1997) suggested a Bayesian bootstrap.[2] Other notable areas of investigation in the L_1-literature concern the study of nonlinear functional forms and structural models with endogeneity [censored quantile regressions: Powell (1984, 1986) Fitzenberger (1997*a*) and Buchinsky and Hahn (1998); simultaneous equations: Amemiya (1982) and Hong and Tamer (2003)]. More recently, authors have allowed for misspecification [Kim and White (2002), Komunjer (2005), Jung (1996)].

In the context of LAD-based inference, 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 heteroskedastic 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 context in Dufour (1981), Campbell and Dufour (1995, 1997) and for L_1-estimation in Dielman and Pfaffenberger (1988*a*, 1988*b*), De Angelis, Hall and Young (1993), Buchinsky (1995). Inference based on signs constitutes an alternative that does not suffer from these shortcomings.

[2]The reader is referred to Buchinsky (1995, 1998), for a review and to Fitzenberger (1997*b*) for a comparison between these methods.

We study here a linear median regression model where the (possibly dependent) disturbance process is assumed to have a null median conditional on some exogenous explanatory variables and its own past. This setup covers non stochastic heteroskedasticity, standard conditional heteroskedasticity (like ARCH, GARCH, stochastic volatility models, ...) as well as other forms of nonlinear dependence. However, linear autocorrelation in the residuals is not allowed. We first treat the problem of inference and show that pivotal statistics based on the signs of the residuals are available for any sample size. Hence, exact inference and exact simultaneous confidence regions on β can be derived using Monte Carlo tests.

For more general processes which may involve stationary ARMA disturbances, these statistics are no longer pivotal. The serial dependence parameters constitute nuisance parameters. However, transforming sign-based statistics with standard HAC methods allows to asymptotically get rid of these nuisance parameters. We thus extend the validity of the Monte Carlo test method. In such cases, we loose the exactness but keep an asymptotic validity. This asymptotic validity requires less assumptions on moments or the shape of the distribution (such as the existence of a density) than usual asymptotic-based inference (such as results for LAD-based estimators). 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.

The paper is organized as follows. In section 2, we present the model and the notations. Section 3 contains general results on exact inference. They are applied to median regressions in section 4. In section 5, we derive confidence intervals at any given confidence level and illustrate the method on a numerical example. Section 6 is dedicated to the asymptotic validity of the finite-sample inference method. In section 7, we give simulation results from comparisons to usual techniques. Section 8 presents illustrative applications: testing the presence of a drift in the Standard and Poor's composite price index series, and testing for β-convergence between levels of per capita output across the U. S. States. Section 9 concludes. The Appendix contains the proofs.

2. Framework

2.1. Model

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

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

where y_t is a dependent variable, $x_t = (x_{t1}, \dots, x_{tp})'$ 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, $y = (y_1, \dots, y_n)' \in \mathbb{R}^n$ will denote the dependent vector, $X = [x_1, \dots, x_n]'$ 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

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

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 $W_1, \ldots, W_t$.

A common assumption – which allows for general forms of dependence – consists in assuming that $\mathbf{u} = \{u_t : t = 1, 2, \ldots\}$ 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 $\mathcal{F}_t = \sigma(W_1, \ldots, W_t), t = 1, 2, \ldots$.

Definition 2.1 MARTINGALE DIFFERENCE. $\mathbf{u}$ *in the adapted sequence* $\mathcal{S}(\mathbf{u}, \mathcal{F})$ *is a martingale difference sequence with respect to* $\{\mathcal{F}_t : t = 1, 2, \ldots\}$ *iff* $\mathsf{E}(u_t|\mathcal{F}_{t-1}) = 0,\ \forall t \geq 1.$

We shall depart from this usual assumption, which requires the existence of the first moments of u_t. Indeed, our aim is to develop a framework which allows for heteroskedasticity of unknown form. From Bahadur and Savage (1956), it is known 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. We call such a structure a *mediangale*, which may be defined conditional on the design matrix X or unconditionally, as follows.

Definition 2.2 STRICT MEDIANGALE. $\mathbf{u}$ *in the adapted sequence* $\mathcal{S}(\mathbf{u}, \mathcal{F})$ *is a strict mediangale with respect to* $\{\mathcal{F}_t : t = 1, 2, \ldots\}$ *iff* $\mathsf{P}[u_1 < 0] = \mathsf{P}[u_1 > 0] = 0.5$ *and*

$$\mathsf{P}[u_t < 0|\mathcal{F}_{t-1}] = \mathsf{P}[u_t > 0|\mathcal{F}_{t-1}] = 0.5,\ \textit{for}\ t > 1. \tag{2.3}$$

Definition 2.3 STRICT 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 strict mediangale conditional on* X *with respect to* $\{\mathcal{F}_t : t = 1, 2, \ldots\}$ *iff* $\mathsf{P}[u_1 < 0|X] = \mathsf{P}[u_1 > 0|X] = 0.5$ *and*

$$\mathsf{P}[u_t < 0|u_1,\, \ldots,\, u_{t-1},\, X] = \mathsf{P}[u_t > 0|u_1,\, \ldots,\, u_{t-1},\, X] = 0.5,\ \textit{for}\ t > 1. \tag{2.4}$$

The above definitions allow u_t to have a discrete distribution, but exclude the presence of a probability mass at zero. This constraint is relaxed in the following definition.

Definition 2.4 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*

$$\mathsf{P}[u_t > 0|u_1, \ldots, u_{t-1}, X] = \mathsf{P}[u_t < 0|u_1, \ldots, u_{t-1}, X], \;\; for\ t > 1. \tag{2.5}$$

The sign operator $s : \mathbb{R} \to \{-1, 0, 1\}$ is defined as $s(a) = \mathbf{1}_{[0,+\infty)}(a) - \mathbf{1}_{(-\infty,0]}(a)$ where $\mathbf{1}_A(a) = 1$ if $a \in A$ and $\mathbf{1}_A(a) = 0$ if $a \notin A$. For convenience, if $u \in \mathbb{R}^n$, we will note $s(u)$, the n-vector composed by the signs of its components.

Stating that $\{u_t : t = 1, 2, \ldots\}$ is a weak mediangale with respect to $\{\mathcal{F}_t : t = 1, 2, \ldots\}$ is equivalent to assuming that $\{s(u_t) : t = 1, 2, \ldots\}$ is a martingale difference with respect to the same sequence of sub-σ algebras $\{\mathcal{F}_t : t = 1, 2, \ldots\}$. However, the weak conditional mediangale concept differs from a martingale difference on the signs because of the conditioning upon the whole process X. Indeed, the reference sequence of sub-σ algebras is usually taken to $\{\mathcal{F}_t = \sigma(W_1, \ldots, W_t) : t = 1, 2, \ldots\}$. Here, the reference sequence is $\{\mathcal{F}_t = \sigma(W_1, \ldots, W_t, X) : t = 1, 2, \ldots\}$. We shall see later that asymptotic inference may be available under weaker assumptions, as a classical martingale difference on signs or more generally mixing conditions on $\{s(u_t), \sigma(W_1, \ldots, W_t) : t = 1, 2, \ldots\}$. However, the conditional mediangale concept allows one to develop exact inference (conditional on X). We have replaced the difference of martingale assumption on the raw process $\mathbf{u}$ by a quasi-similar hypothesis on a robust transform of this process $\mathbf{s}(\mathbf{u})$. Below we will see it is relatively easy to deal with a weak mediangale by a simple transformation of the sign operator, but 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* $u = (u_1, \ldots, u_n)'$ *satisfy a strict mediangale conditional on* X.

It is easy to see that Assumption 2.1 entails $\mathrm{med}(u_1|x_1, \ldots, x_n) = 0$, and

$$\mathrm{med}(u_t|x_1, \ldots, x_n, u_1, \ldots, u_{t-1}) = 0, \;\; t = 2, \ldots, n. \tag{2.6}$$

Hence, we are in a median regression context. Our last 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 $\mathcal{F}_t = \sigma(W_1, \ldots, W_t)$, $t = 1, 2, \ldots$.[3]

[3] X is strongly exogenous for β if X is sequentially exogenous and if Y does not Granger cause X; see Gouriéroux and Monfort (1995, Volume 1).

Proposition 2.1 MEDIANGALE EXOGENEITY. Suppose $\{x_t : t = 1, 2, \dots\}$ 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, heteroskedastic or dependent ones, as long as conditional medians are 0. We stress that 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) and Campbell and Dufour (1991, 1995, 1997).[4]

Asymptotic normality of the LAD estimator, which is presented in its most general way in Fitzenberger (1997*b*), holds under some mixing concepts on $\{s(u_t), \sigma(W_1, \dots, W_t) : t = 1, 2, \dots\}$ and an orthogonality condition between $s(u_t)$ and x_t. Besides, it requires additional assumptions on moments.[5] 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.

2.2. Special cases

The above framework obviously covers independence but also a large spectrum of heteroskedasticity and dependence patterns. For example, it 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: (1) ARCH(q) with non-Gaussian

[4] Assumption **2.2** 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 $\mathsf{P}[u_t < 0|u_1, \dots, u_{t-1}, x_1, \dots, x_t] = q$ in Proposition 2.1. However, with error heterogeneity 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. A classical result in nonparametric 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, Chapter 2). For a recent econometric exploitation of this result, see Chernozhukov, Hansen and Jansson (2006). Proposition 2.1 above provides general conditions under which such a result holds for non-*i.i.d.* observations.

[5] Fitzenberger (1997*b*) 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.

noises ε_t's,

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

(2) GARCH(p, q) with non-Gaussian noises ε_t's,

$$\sigma_t(x_1, \ldots, x_{t-1}, u_1, \ldots, u_{t-1})^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \cdots + \alpha_q u_{t-q}^2 + \gamma_1 \sigma_{t-1}^2 + \cdots + \gamma_p \sigma_{t-p}^2 ;$$

(3) stochastic volatility models with non-Gaussian noises ε_t's,

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

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. For nonstationary processes that satisfy the mediangale assumption, sign-based inference will work whereas all inference procedures based on asymptotic behavior of estimators may fail or require difficult validity proofs.

3. Exact finite-sample sign-based inference

The most common procedure for developing inference on a statistical model can be described as follows. First, one finds a (hopefully consistent) estimator; second, the asymptotic distribution of the latter is established, from which confidence sets and tests are derived. Here, we shall proceed in the reverse order. We study first the test problem, then build confidence sets, and finally estimators.[6] Hence, results on the valid finite-sample test problem will be adapted to obtain valid confidence intervals and estimators.

3.1. Motivation

In econometrics, tests are often based on t or χ^2 -statistics, which are derived from asymptotically normal statistics with a consistent estimator of the asymptotic covariance matrix. Unfortunately, in finite samples, these first-order approximations can be misleading. Test sizes can be quite far from their nominal level: both the probability that an asymptotic test rejects a correct null hypothesis and the probability that a component of β is contained in an asymptotic confidence interval may differ considerably from assigned nominal levels. One can find examples of such distortions in the dynamic literature [see for example Dufour

[6] For the estimation theory, the reader is referred to Coudin and Dufour (2007).

(1981), Mankiw and Shapiro (1986) and Campbell and Dufour (1995, 1997)]; on inference based on L_1-estimators [see Dielman and Pfaffenberger (1988*a*, 1988*b*), Buchinsky (1995) and 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 asymptotics and yields no guarantee that the level condition be satisfied in finite samples.

Another way to appreciate the unreliability of asymptotic methods in finite samples is to recall the theorem established by Lehmann and Stein (1949). Consider testing whether n observations are independent with common zero median:

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

Testing H_0 turns to check whether the joint distribution F_n of the observations belongs to the set $\mathcal{H}_0 = \{F_n \in \mathcal{F}_n : F_n \text{ satisfies } H_0\}$ without any other restriction. In other words, H_0 allows for heteroskedasticity of unknown form. For this setup, Lehmann and Stein (1949) established the following theorem [see also Pratt and Gibbons (1981, Chap. 4, Sect. 3, p. 218)].

Theorem 3.1 *If a test has level α for H_0, where $0 \leq \alpha < 1$, then it must satisfy* $\mathsf{P}[\,\textit{Rejecting } H_0 \mid |X_1|, \ldots, |X_n|\,] \leq \alpha$ *under* H_0 .

The level of a valid test must equal α conditional on the observation absolute values. Theorem **3.1** also implies that any procedure that does not satisfy the above condition has size one. It is not clear that least square-based procedures typically designated as "robust to heteroskedasticity" or "HAC" [see White (1980), Newey and West (1987), Andrews (1991), etc.] do satisfy Theorem **3.1** condition. For some examples of size distortion in some specific setups, see the simulation study in section 7.

Sign-based procedures do satisfy this condition. Besides, as we will show in section 4, distribution-free sign-based statistics are available even in finite samples. They have been used in the statistical literature to derive nonparametric sign tests. The combination of both remarks give the theoretical basis for developing an exact inference method.

3.2. Distribution-free pivotal functions and nonparametric tests

When the disturbance process is a conditional mediangale, the joint distribution of the signs of the disturbances is completely determined. These signs are *i.i.d.* and take the values 1 and -1 with equal probability $1/2$. This result is stated more precisely in the following proposition. The case with a mass at zero can be covered provided a transformation in the sign operator definition.

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

$$\mathsf{P}[s(u_t) = 1 \,|x_1, \dots, x_n] = \mathsf{P}[s(u_t) = -1 \,|x_1, \dots, x_n] = \frac{1}{2}, \quad t = 1, \dots, n. \tag{3.2}$$

More generally, this result holds for any combination of $t = 1, \dots, n$. 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 **3.2**, it follows that the residual sign vector

$$s(y - X\beta) = [s(y_1 - x_1'\beta), \dots, s(y_n - x_n'\beta)]' \tag{3.3}$$

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 $T = T\big(s(y - X\beta), X\big)$ 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 **3.2**, 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 $H_0(\beta_0)$, $s(y_t - x_t'\beta_0) = s(u_t),\ t = 1, \dots, n$. Thus, conditional on X,

$$T\big(s(y - X\beta_0), X\big) \sim T(S_n, X) \tag{3.4}$$

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

$$T\big(s(y - X\beta_0), X\big) > c_T(X, \alpha) \tag{3.5}$$

where $c_T(X, \alpha)$ is the $(1-\alpha)$-quantile of the distribution of $T(S_n, X)$. This extends results in Dufour (1981) and Campbell and Dufour (1991, 1995, 1997).

This method can be extended to error distributions with a mass at zero, $i.e.$, $\mathsf{P}[u_t = 0 \,|\, X, u_1, \dots, u_{t-1}] = p_t(X, u_1, \dots, u_{t-1}) > 0$ where the $p_t(\cdot)$ are unknown and may vary between observations. A way out consists in modifying the sign function $s(x)$ as $\tilde{s}(x, V) = s(x) + \big[1 - s(x)^2\big] s(V - 0.5)$, where $V \sim \mathcal{U}(0, 1)$. If V_t is independent of u_t then, irrespective of the distribution of u_t,

$$\mathsf{P}[\tilde{s}(u_t, V_t) = +1] = \mathsf{P}[\tilde{s}(u_t, V_t) = -1] = \frac{1}{2}. \tag{3.6}$$

This yields the following proposition.

Proposition 3.3 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* $\tilde{s}_t = \tilde{s}(u_t, V_t)$ *are* i.i.d. *conditional on* X *with the distribution*

$$\mathsf{P}[\tilde{s}_t = 1 \,|\, X] = \mathsf{P}[\tilde{s}_t = -1 \,|\, X] = \frac{1}{2}, \quad t = 1, \ldots, n. \tag{3.7}$$

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 $T\big(\tilde{s}(y - X\beta_0, X)\big)$ has to be used instead of $T\big(s(y - X\beta_0, X)\big)$.

4. Regression sign-based tests

In this section, we present sign-based test statistics that are pivots and provide power against alternatives 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.

4.1. Regression sign-based statistics

The class of pivotal functions studied in the previous section is quite general. So, we wish to choose a test statistic (the form of the T function) that can have power against alternatives of interest. Unfortunately, there is no uniformly most powerful test of $H_0(\beta_0) : \beta = \beta_0$ against $H_1(\beta_0) : \beta \neq \beta_0$. Hence, for testing $H_0(\beta_0)$ against $H_1(\beta_0)$ in model (2.1), we consider test statistics of the following form:

$$D_S(\beta_0,\, \varOmega_n) = s(y - X\beta_0)' X \varOmega_n\big(s(y - X\beta_0),\, X\big) X' s(y - X\beta_0) \tag{4.1}$$

where $\varOmega_n\big(s(y - X\beta_0), X\big)$ is a $p \times p$ weight matrix that depends on the *constrained* signs $s(y - X\beta_0)$ under $H_0(\beta_0)$. The latter feature of the weight matrix allows one to obtain a finite-sample distributional theory for $D_S(\beta_0, \varOmega_n)$. The weighting matrix $\varOmega_n\big(s(y - X\beta_0), X\big)$ 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, $\varOmega_n\big(s(y - X\beta_0),\, X\big)$ 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].[7]

[7]Under more restrictive assumptions, statistics which exploit other robust functions of $y - X\beta_0$ [such as ranks or signed ranks] can lead to more powerful tests. However, the fact we allow for both heteroskedasticity and nonlinear 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 setup. For discussion of such alternative statistics (applicable under stronger assumptions), see Hallin and Puri (1991, 1992), Hallin, Vermandele and Werker (2006, 2008), Hallin and Werker (2003) and the references therein.

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

$$SB(\beta_0) = s(y - X\beta_0)' X X' s(y - X\beta_0) = \|X' s(y - X\beta_0)\|^2 , \tag{4.2}$$

$$SF(\beta_0) = s(y - X\beta_0)' P(X) s(y - X\beta_0) = \|X' s(y - X\beta_0)\|_M^2 , \tag{4.3}$$

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 $f'(0) = 0$].[8] Their proof can easily be extended to disturbances that satisfy the mediangale property and for which the conditional density at zero is the same $f_t(0|X) = f(0|X), \ \forall t.$

$SF(\beta_0)$ can be interpreted as a sign analogue of the Fisher statistic. More precisely, $SF(\beta_0)$ is a monotonic transformation of the Fisher statistic for testing $\gamma = 0$ in the regression of $s(y - X\beta_0)$ on X: $s(y - X\beta_0) = X\gamma + v$. This remark holds also for a general sign-based statistic of the form 4.1, 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 assume *i.i.d.* errors and a fixed design matrix. In that setup, 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. It does not require one to estimate the density of the disturbance at zero contrary to LR and Wald-type statistics.

The Boldin et al. (1997) local optimality interpretation can be extended to heteroskedastic 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 4.1 *In model* (2.1)*, suppose the mediangale Assumption 2.1 holds, and the disturbances are heteroskedastic with conditional densities* $f_t(\,\cdot\,|X), \ t = 1, 2, \ldots,$ *that 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*

$$\tilde{SB}(\beta_0) = s(y - X\beta_0)' \tilde{X} \tilde{X}' s(y - X\beta_0) \tag{4.4}$$

[8] 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, Fomenko and Novikov (1984, Ch. 2, pp. 76-86) or Gray (1998, Ch. 21, pp. 373-380) for a discussion of various curvature notions.

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

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 (4.1) 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. We saw in the first section that this property is induced by the mediangale Assumption 2.1. 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 heteroskedastic disturbances, the set of optimal estimating equations is $\sum_{t=1}^{n} s(y_t - x_t'\beta) \otimes \tilde{x}_t = 0$. In those cases, X (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 $J_n\big(s(y - X\beta_0), X\big)$. We consider $\Omega_n\big(s(y - X\beta_0), X\big) = \frac{1}{n}\hat{J}_n\big(s(y - X\beta_0), X\big)^{-1}$ where $\hat{J}_n\big(s(y - X\beta_0), X\big)$ stands for a consistent estimate of $J_n\big(s(y - X\beta_0), X\big)$ that can be obtained using kernel-estimators, for example [see Parzen (1957), White (2001), Newey and West (1987), Andrews (1991)]. This leads to

$$D_S(\beta_0, \frac{1}{n}\hat{J}_n^{-1}) = \frac{1}{n} s(y - X\beta_0)' X \hat{J}_n^{-1} X' s(y - X\beta_0). \tag{4.5}$$

$J_n\big(s(y - X\beta_0), X\big)$ 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.

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. 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.

4.2. 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). It proceeds as follows. Let us consider a statistic T whose conditional distribution given X is continuous and free of nuisance parameters, and a test which rejects the null hypothesis when $T \geq c(\alpha)$. We denote by $G(x) = P[T \geq x]$ the survival function, and by $F(x) = P[T \leq x]$ the distribution function. Let $T^{(0)}$ be the observed value of T, and $T^{(1)}, \ldots, T^{(N)}$, N independent replicates of T. The empirical p-value is given by

$$\hat{p}_N(x) = \frac{N\hat{G}_N(x) + 1}{N + 1} \tag{4.6}$$

where $\hat{G}_N(x) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{[0,\infty)}(T^{(i)} - x)$. Then we have

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

where $I[x]$ stands for the largest integer less than equal to x; see Dufour (2006). If N is such that $\alpha(N+1) \in \mathbb{N}$, then $\mathsf{P}[\hat{p}_N(T^{(0)}) \leq \alpha] = \alpha$: the test level is exactly controlled.

In the case of **discrete distributions**, the method must be adapted to deal with ties. Indeed, the usual order relation on $\mathbb{R}$ is not appropriate for comparing discrete realizations that have a strictly positive probability to be equal.[9] Here, we use a randomized tie-breaking procedure for evaluating empirical survival functions. The latter is based on replacing the usual order relation by a lexicographic order relation [see Dufour (2006)]. 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:

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

This leads to the following p-value function:

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

[9] Different procedures have been presented in the literature. They can be classified between randomized and nonrandomized procedures, both aiming to exactly control back the level of the test. For a good review of this problem, the reader is referred to Coakley and Heise (1996).

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

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

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

To illustrate the method, consider testing $H_0(\beta_0)$ in (2.1) under a mediangale assumption on the errors, and using $D_S(\beta, X'X^{-1})$. After computing $SF^{(0)} = D_S(\beta_0, X'X^{-1})$ from the data, choose N the number of replicates, such that $\alpha(N+1)$ is an integer, where α is the desired level. Then, generate N replicates $SF^{(j)} = S^{(j)\prime}X(X'X)^{-1}X'S^{(j)}$ where $S^{(j)}$ is a realization of a n-vector of independent uniform Bernoulli random variables, and compute $\tilde{p}_N(SF^{(0)})$. Finally, the Monte Carlo test rejects $H_0(\beta_0)$ with level α if $\tilde{p}_N(SF^{(0)}) < \alpha$.

5. Regression sign-based confidence sets

In the previous section, we have shown how to obtain Monte Carlo sign-based joint tests for which we can exactly control the level, for any given finite number of observations. In this section, we discuss how to use such tests in order to build confidence sets for β with known level. 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 nonlinear ones. Obviously, obtaining a continuous grid of $\mathbb{R}^p$ is not realistic. We will instead require **global optimization search algorithms**.

5.1. Confidence sets and conservative confidence intervals

Projection techniques yield finite-sample valid confidence intervals and confidence sets for general functions of the parameter β.[10] The basic idea is the following one. Suppose a simultaneous confidence set with level $1-\alpha$ for β, $C_{1-\alpha}(\beta)$, is available. Since $\beta \in C_{1-\alpha}(\beta) \Rightarrow g(\beta) \in g\big(C_{1-\alpha}(\beta)\big)$, we have $\mathsf{P}[\beta \in C_{1-\alpha}(\beta)] \geq 1-\alpha \Rightarrow \mathsf{P}[g(\beta) \in g\big(C_{1-\alpha}(\beta)\big)] \geq 1-\alpha$. Thus, $g\big(C_{1-\alpha}(\beta)\big)$ is a conservative confidence set

[10]For examples of use in different settings and for further discussion, see Dufour (1990, 1997), Abdelkhalek and Dufour (1998), Dufour and Kiviet (1998), Dufour and Jasiak (2001), Dufour and Taamouti (2005).

for $g(\beta)$. If $g(\beta)$ is scalar, the interval (in the extended real numbers) $I_g[C_{1-\alpha}(\beta)] = \big[\inf_{\beta \in C_{1-\alpha}(\beta)} g(\beta)\,,\ \sup_{\beta \in C_{1-\alpha}(\beta)} g(\beta)\big]$ 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". The optimization problems are stated here for the statistic SF:

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

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

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

5.2. Numerical illustration

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

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

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 2-vector, we can provide a graphical illustration. To each value of the vector β is associated the corresponding simulated p-value. Confidence 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 nonelliptic 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

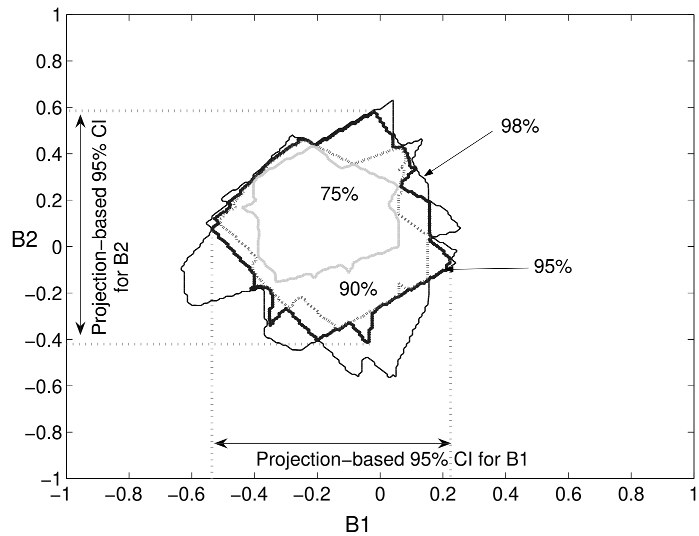


Figure 1. Confidence regions provided by SF-based inference

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]

outliers.

6. 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.

6.1. 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 (4.5) allows one to obtain an asymptotically pivotal function. The set of assumptions we make to stabilize the asymptotic behavior will be needed for further asymptotic results. We consider the linear model (2.1), with the following assumptions.

Assumption 6.1 MIXING. *$\{(x_t', u_t) : t = 1, 2, \ldots\}$ is α-mixing of size $-r/(r-2)$, $r > 2$.*[11]

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

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

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

Assumption 6.5 CONSISTENT ESTIMATOR OF J_n. *$\varOmega_n(\beta_0)$ is symmetric positive definite uniformly over n and $\varOmega_n - \frac{1}{n}J_n^{-1} \xrightarrow{p} 0$.*

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

Theorem 6.1 ASYMPTOTIC DISTRIBUTION OF SIGN-BASED STATISTICS. *In model (2.1), with Assumptions 6.1- 6.5, we have, under $H_0(\beta_0)$, $D_S(\beta_0, \varOmega_n) \to \chi^2(p)$.*

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

Corollary 6.2 *In model (2.1), suppose the mediangale Assumption 2.1 and boundedness Assumption 6.3 are fulfilled. If $X'X/n$ is positive definite uniformly over n and converges in probability to a definite positive matrix, then, under $H_0(\beta_0)$, $SF(\beta_0) \to \chi^2(p)$.*

6.2. Asymptotic validity of Monte Carlo tests

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

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

6.2.1. Generalities

Let us consider a parametric or semiparametric model $\{\mathrm{M}_\beta,\ \beta \in \Theta\}$. 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 $F_n(x)$. 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 **6.3**, we show that if a sequence of conditional survival functions $\tilde{G}_n\big(x|X_n(\omega)\big)$ given $X(\omega)$ satisfies $\tilde{G}_n\big(x|X_n(\omega)\big) \to G(x)$ with probability one, where G does not depend on the realization $X(\omega)$, then $\tilde{G}_n\big(x|X_n(\omega)\big)$ 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 6.3 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 $\tilde{G}_n\big(x|X_n(\omega)\big) \underset{n\to\infty}{\longrightarrow} G(x)$ with probability one. Then

$$\lim_{n\to\infty} \mathsf{P}[\tilde{G}_n(c_n S_n(\beta_0),\ X_n(\omega)) \leq \alpha] \leq \alpha. \tag{6.1}$$

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 resp. 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

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

$\tilde{G}_n^{(N)}\big(x, n, S_n^{(0)}, S_n(N), U(N+1)\big)$ 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

$$\tilde{p}_n^{(N)}(x) = \frac{N\tilde{G}_n^{(N)}(x) + 1}{N + 1}. \tag{6.2}$$

We can now state the Monte Carlo-based version of Theorem **6.3**.

Theorem 6.4 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 **6.3** and **6.4** are much more general. They do not exclusively rely on asymptotic normality: the limiting distribution may be different from a Gaussian one. Besides, the rate of convergence may differ from $\sqrt{n}$.

6.2.2. Asymptotic validity of sign-based inference

In model (2.1), suppose that conditions 6.1- 6.5 hold and consider the testing problem: $H_0(\beta_0) : \beta = \beta_0$ against $H_1(\beta_0) : \beta \neq \beta_0$. Let $D_S(\beta, \hat{J}_n^{-1})$ be the test statistic as defined in (4.5). 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 $B(1, .5)$ 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 (6.2). From Theorem **6.4**, 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 behavior of estimators and consequently is more restrictive. More moments existence restrictions are needed, see Fitzenberger (1997*b*) and Weiss (1991). Besides, asymptotic variance of the LAD estimator involves the conditional density at zero of the disturbance process $\{u_t : \ t = 1, 2, \dots\}$ 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.

7. Simulation study

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

$$y_t = x_t'\beta_0 + u_t, \ \ t = 1, \dots , n, \tag{7.1}$$

where $x_t = (1, x_{2,t}, x_{3,t})'$ and β_0 are 3×1 vectors. We denote the sample size n. We investigate the behavior of inference and confidence regions for 13 general DGP's that are presented in Table 2. For the first 7 ones, $\{u_t : \ t = 1, 2 \dots\}$ is *i.i.d.* or depends on the explanatory variables and its past values in a *multiplicative* heteroskedastic or dependent and stationary way, $u_t = h(x_t, u_{t-1}, \dots , u_1)\epsilon_t, \ t = 1, \dots , n$. 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. Next, we study the behavior of the sign-based inference (involving a HAC correction) when inference is only asymptotically valid. In cases C8-C10, x_t and u_t are such that $\mathsf{E}(u_t x_t) = 0$ and $\mathsf{E}[s(u_t)x_t] = 0$ for all t. Finally, cases C11 and C12 illustrate two kinds of second-order nonstationary disturbances. As we noted previously, sign-based inference does not require stationary assumptions in contrast with tests derived from CLT.

Cases C1 and C2 present *i.i.d.* normal observations without and with conditional heteroskedasticity. Case C3 involves outliers in the error term. This can be seen as an example of measurement error in the observed y. Cases C4 and C5 involve other heteroskedastic schemes with stationary GARCH and stochastic volatility disturbances. Case C6 is a very unbalanced design matrix (where the LAD estimator performs poorly). Case C6 BIS combines the previous unbalanced scheme in the design matrix with heteroskedastic disturbances. Case C7 is an example of heavy-tailed errors (Cauchy). Cases C8, C9

Table 2. Simulated models

C 1:	Normal HOM:	$(x_{2,t}, x_{3,t}, u_t)' \overset{i.i.d}{\sim} \mathcal{N}(0, I_3),\ t = 1, \ldots, n$
C 2:	Normal HET:	$(x_{2,t}, x_{3,t}, \tilde{u}_t)' \overset{i.i.d}{\sim} \mathcal{N}(0, I_3)$ $u_t = min\{3, max[0.21, \lvert x_{2,t}\rvert]\} \times \tilde{u}_t,\ t = 1, \ldots, n$
C 3:	Outlier:	$(x_{2,t}, x_{3,t})' \overset{i.i.d.}{\sim} \mathcal{N}(0, I_2)$, $u_t \overset{i.i.d.}{\sim} \begin{cases} N[0,\ 1] & \text{with p} = 0.95 \\ N[0,\ 1000^2] & \text{with p} = 0.05 \end{cases}$ $x_t, u_t,$ independent, $t = 1, \ldots, n.$
C 4:	Stat. GARCH(1,1):	$(x_{2,t}, x_{3,t})' \overset{i.i.d.}{\sim} \mathcal{N}(0, I_2),\ u_t = \sigma_t \epsilon_t$ with $\sigma_t^2 = 0.666u_{t-1}^2 + 0.333\sigma_{t-1}^2$ where $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$, $x_t, \epsilon_t,$ independent, $t = 1, \ldots, n.$
C 5:	Stoc. Volatility:	$(x_{2,t}, x_{3,t})' \overset{i.i.d.}{\sim} \mathcal{N}(0, I_2), u_t = exp(w_t/2)\epsilon_t$ with $w_t = 0.5w_{t-1} + v_t$, where $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$, $v_t \overset{i.i.d.}{\sim} \chi_2(3)$, $x_t, u_t,$ independent, $t = 1, \ldots, n.$
C 6:	Deb. design mat.:	$x_{2,t} \sim \mathcal{B}(1, 0.3),\ x_{3,t} \overset{i.i.d.}{\sim} \mathcal{N}(0, .01^2)$, $u_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$, x_t, u_t independent, $t = 1, \ldots, n.$
C 6 BIS:	Deb. design matrix + HET. dist.:	$x_{2t} \overset{i.i.d}{\sim} \mathcal{N}(0, 1)$, $x_{3t} \overset{i.i.d}{\sim} \chi_2(1)$, $u_t = x_{3t}\epsilon_t,\ \epsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0, 1),\ x_t, \epsilon_t$ independent, $t = 1, \ldots, n.$
C 7:	Cauchy disturbances:	$(x_{2,t}, x_{3,t})' \sim \mathcal{N}(0, I_2)$, $u_t \overset{i.i.d.}{\sim} \mathcal{C}$,$x_t, u_t,$ independent, $t = 1, \ldots, n.$
C 8:	AR(1)-HOM, $\rho_u = .5$:	$(x_{2,t}, x_{3,t}, \nu_t^u)' \sim \mathcal{N}(0, I_3), t = 2, \ldots, n,$ $u_t = \rho_u u_{t-1} + \nu_t^u$, $(x_{2,1}, x_{3,1})' \sim \mathcal{N}(0, I_2)$, ν_1^u insures stationarity.
C 9:	AR(1)-HET, $\rho_u = .5,$: $\rho_x = .5$	$x_{j,t} = \rho_x x_{j,t-1} + \nu_t^j,\ j = 1, 2,$ $u_t = min\{3, max[0.21, \lvert x_{2,t}\rvert]\} \times \tilde{u}_t$, $\tilde{u}_t = \rho_u \tilde{u}_{t-1} + \nu_t^u$, $(\nu_t^2, \nu_t^3, \nu_t^u)' \overset{i.i.d}{\sim} \mathcal{N}(0, I_3),\ t = 2, \ldots, n$ ν_1^2, ν_1^3 and ν_1^u chosen to insure stationarity.
C 10:	AR(1)-HOM, $\rho_u = .9$:	$(x_{2,t}, x_{3,t}, \nu_t^u)' \sim \mathcal{N}(0, I_3), t = 2, \ldots, n,$ $u_t = \rho_u u_{t-1} + \nu_t^u$, $(x_{2,1}, x_{3,1})' \sim \mathcal{N}(0, I_2)$, ν_1^u insures stationarity.
C 11:	Nonstat. GARCH(1,1):	$(x_{2,t}, x_{3,t}, \epsilon_t)' \overset{i.i.d.}{\sim} \mathcal{N}(0, I_3),\ t = 1, \ldots, n,$ $u_t = \sigma_t \epsilon_t,\ \sigma_t^2 = 0.8u_{t-1}^2 + 0.8\sigma_{t-1}^2.$
C 12:	Exp. Var.:	$(x_{2,t}, x_{3,t}, \epsilon_t)' \overset{i.i.d.}{\sim} \mathcal{N}(0, I_3), u_t = exp(.2t)\epsilon_t.$

and C10 illustrate the behavior of sign-based inference when the error term involves linear dependence at different levels. Finally, cases C11 and C12 involve disturbances that are not second-order stationary (nonstationary GARCH and exponential variance) but for which the mediangale assumption holds. The design matrix is simulated once for all the presented cases. Hence, results are conditional. Cases C1-2, C8-10 have been used by Fitzenberger (1997*b*) to study the performance of block bootstrap (MBB).

7.1. 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 $SF = D_S\big(\beta,(X'X)^{-1}\big)$ 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 3 different covariance estimators: the usual under homoskedasticity and independence (IID), White correction for heteroskedasticity (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),[12] Bartlett kernel covariance estimator with automatic bandwidth parameter [$BT,$ Powell (1984), Buchinsky (1995)], design matrix bootstrap centering around the sample estimate [$DMB,$ Buchinsky (1998)], moving block bootstrap centering around the sample estimate [MBB, Fitzenberger (1997*b*)].[13] 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, LM) of asymptotic tests are compared in our setup.

In Tables 3 and 4, 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.

Table 3 contains models when the mediangale condition 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 heteroskedasticity (cases C4, C6 BIS). The asymptotic versions of sign-based tests suffer from the same under-rejection than other asymptotic tests, suggesting that, for small samples ($n = 50$), the distribution of the test statistic is

[12] 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)].

[13] The block size is 5.

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										
C 1: HOM	.052	.050	.086	.050	.089	.047	.068	.060	.096	.113
$\rho_\epsilon = \rho_x = 0,$	*.047**	*.019**								
C 2: HET	.052	.057	.300	.037	.059	.051	.137	.162	.100	.118
$\rho_\epsilon = \rho_x = 0,$	*.045**	*.023**								
C 3:	.047	.048	.088	.043	.083	.039	.066	.056	.008	.009
Outlier	*.044**	*.015**								
C 4:	.042	.046	.040	.005	.005	.004	.012	.080	.046	.046
St. GARCH(1,1)	*.040**	*.013**								
C 5:	.043	.041	.063	.006	.014	.006	.031	.054	.014	.014
Stoch. Volat.	*.045**	*.021**								
C 6:	.047	.049	.080	.048	.084	.043	.064	.085	.060	.095
Debalanced	*.043**	*.022**								
C 7:	.058	.059	.069	.013	.033	.012	.044	.061	.023	.023
Cauchy	*.049**	*.021**								
Nonstationary models										
C 6 BIS:	.044	.042	.687	.020	.044	.152	.307	.421	.171	.173
Deb.+ Het.	*.040**	*.018**								
C 11: Nonst.	.054	.057	.003	.000	.001	.000	.002	.060	.016	.016
GARCH(1,1)										
C 12: Exp. Var.	.049	.051	.017	.000	.000	.000	.000	.132	.014	.014

* Sizes using asymptotic critical values based on $\chi^2(3)$.

really far from its asymptotic limit. Hence, the sign-based method that deals directly with this distribution has clearly an advantage on asymptotic methods. When the dependence in the disturbance process is highly nonlinear (case C6 BIS), the kernel estimation of the LAD asymptotic covariance matrix is not reliable anymore.

In Table 4, we illustrate behaviors 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 to 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 under-rejection. There exists 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 over-rejections of asymptotic Wald tests based on HAC estimation of the asymptotic covariance matrix when the data set involves a small number of observations. These results suggest, in a sense, that when the data suffer from both

Table 4. Linear regression with serial dependence: 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
Serial dependence (cases when mediangale condition fails)										
C 8: HOM	.126	.022	.171	.124	.118	.085	.151	.201	.240	212
$\rho_\epsilon = .5,\ \rho_x = 0$	-	*.019**								
C 9: HET	.218	.026	.440	.131	.097	.108	.308	.407	.328	.276
$\rho_\epsilon = \rho_x = .5$	-	*.017**								
C 10: HOM	.521	.012	.553	.516	.339	.355	.551	.649	.677	.534
$\rho_\epsilon = .9,\ \rho_x = 0$**	-	*.003**								

$*$ Sizes using asymptotic critical values based on $\chi^2(3)$.

$**$ Automatic bandwidth parameters are restricted to be < 10 to avoid invertibility problems.

a small number of observations and linear dependence, the first problem to solve is the finite-sample distortion, which is not what is usually done.

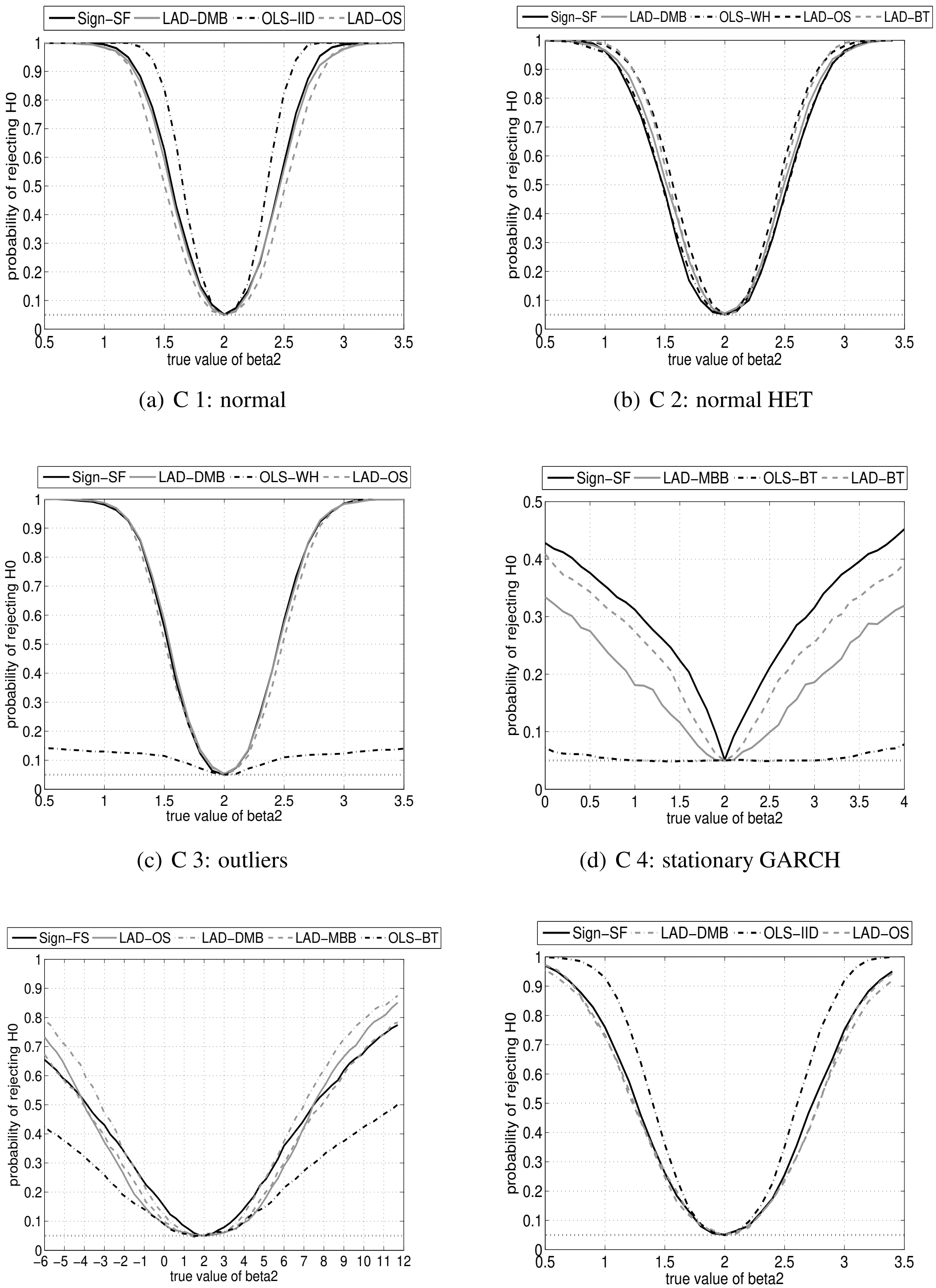
7.2. Power

Then, we illustrate the **power** of these tests. We are particularly interested in comparing the sign-based inference to kernel and bootstrap methods. 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, loose power.

Sign-based inference has a comparable power performance with usual methods in cases C1, C2, C3, C6, C9 with the advantage that the level is exactly controlled, which leads to great difference in small samples. In heteroskedastic or heterogenous cases (C4, C5, C7, C11, C12), sign-based inference greatly dominates other methods: levels are exactly controlled and power functions largely 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 $D_S\big(\beta, \hat{J}_n^{-1}\big)$, which is still asymptotically valid, seems to lead to good power performance for a mild autocorrelation, along with a better size control (cases C9 and C10).[14] Only for very high autocorrelation (close to unit root process), the sign-based inference is not adapted.

[14]The power functions for case C8 are not reported here as they lead to similar results as case C9.

Figure 2. Power functions (1)



(a) C 1: normal

(b) C 2: normal HET

(c) C 3: outliers

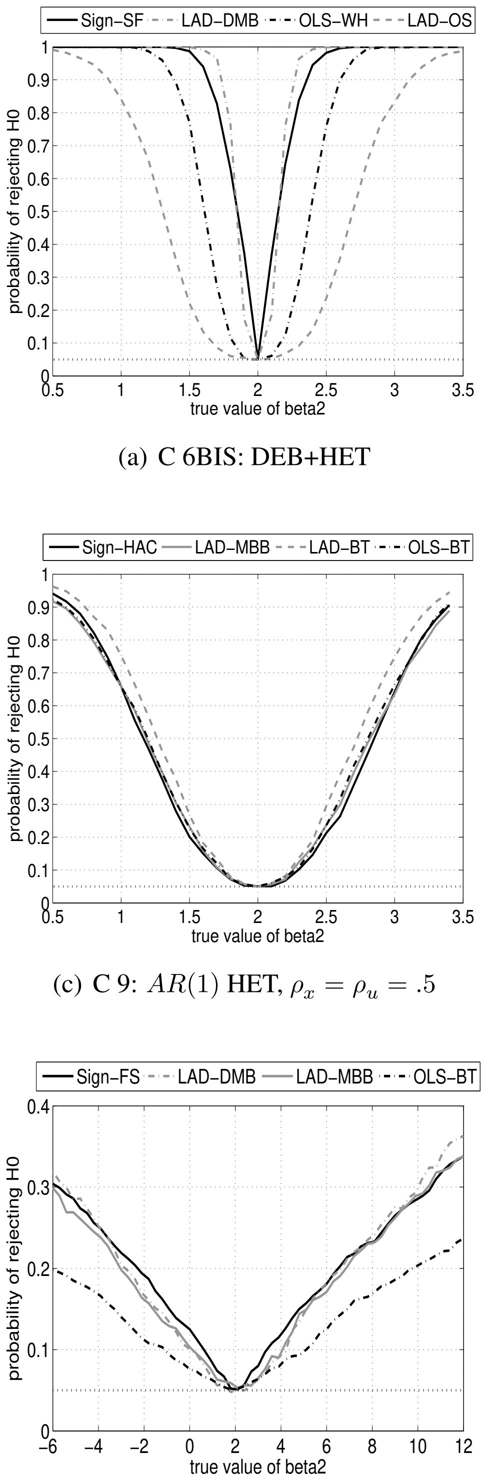
(d) C 4: stationary GARCH

(e) C 5: stochastic volatility

(f) C 6: DEB

Powers are level corrected. Sign: $SF = D_S(\beta, X'X^{-1})$, $SHAC = D_S(\beta, \hat{J}_n^{-1})$; LAD/OLS: $DMB =$ design matrix boot., $MBB =$ moving block boot.; $BT =$ Bartlett kern., $IID =$ homo., $WH =$ White cor., $OS =$ order stat.

Figure 3. Power functions (2)



(a) C 6BIS: DEB+HET

(b) C 7: Cauchy

(c) C 9: $AR(1)$ HET, $\rho_x = \rho_u = .5$

(d) C 10: $AR(1)$ HOM, $\rho_u = .9$

(e) C 11 : Nonstationary GARCH

(f) C 12: exponential variance

Powers are level corrected. Sign: $SF = D_S(\beta, X'X^{-1})$, $SHAC = D_S(\beta, \hat{J}_n^{-1})$; LAD/OLS: $DMB =$ design matrix boot., $MBB =$ moving block boot.; $BT =$ Bartlett kern., $IID =$ homo., $WH =$ White cor., $OS =$ order stat.

7.3. Confidence intervals

As the sign-based confidence regions are by construction of level higher that $1 - \alpha$ 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 5. We only consider the stationary examples. In the nonstationary cases, inference based on t-statistics may not mean anything. 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.

8. Examples

In this section, two illustrative applications of the sign-based inference are presented. One on financial data, one in growth theory. First, we consider testing a drift on the Standard and Poor's composite price index (S&P) 1928-1987, which is known to involve a large amount of heteroskedasticity. We consider robust tests on the whole period and on the 1929 Krach subperiod. In the second illustration, we test for the presence of β-convergence across the U.S. States during the 1880-1988 period using the Barro and Sala-i-Martin (1991) data set. Finite-sample sign-based inference is also particularly adapted to regional data sets, which have by nature fixed sample size.

8.1. Standard and Poor's drift

We test the presence of a drift on the Standard and Poor's composite price index (SP), 1928-1987. That process is known to involve a large amount of heteroskedasticity and have been used by Gallant, Hsieh and Tauchen (1997) and Dufour and Valéry (2008) to fit a stochastic volatility model. Here, we are interested in robust testing without modeling 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 model with a constant and a drift:

$$y_t = a + bt + u_t, \quad t = 1, \ldots, 16127, \tag{8.1}$$

where we let the possibility that $\{u_t : t = 1, \ldots, 16127\}$ presents a stochastic volatility or any kind of nonlinear heteroskedasticity of unknown form. White and Breusch-Pagan tests for heteroskedasticity both reject homoskedasticity at 1%.[15]

[15] White: 499 (p-value=.000) ; BP: 2781 (p-value=.000).

Table 5. 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														
C 1:	av. spread	1.29	1.52	1.40	1.16	1.36	1.02	.81	.90	.89	.79	.88	.85	.82	.88	.87
$\rho_u = \rho_x = 0$	(st. dev.)	(.21)	(.27)	(.29)	(.14)	(.28)	(.29)	(.23)	(.21)	(.22)	(.21)	(.24)	(.24)	(.15)	(.19)	(.22)
HOM	cov. lev.	1.0	1.0	1.0	1.0	1.0	1.0	.97	.97	.97	.95	.96	.95	.97	.96	.96
C 2:		.76	1.43	.74	.66	1.26	.48	.43	.94	.43	.42	.90	.41	.50	.92	.50
$\rho_u = \rho_x = 0$		(.14)	(.29)	(.17)	(.15)	(.28)	(.18)	(.09)	(.24)	(.11)	(.10)	(.27)	(.12)	(.11)	(.29)	(.11)
HET		1.0	1.0	1.0	1.0	1.0	1.0	.98	.97	.99	.97	.95	.97	.99	.95	.99
C 3:		1.26	1.37	1.05	1.15	1.24	.91	.92	.94	.98	.88	.98	1.04	.88	.88	.88
Outlier		(.26)	(.31)	(.30)	(.25)	(.29)	(.30)	(.80)	(.79)	(1.29)	(.67)	(1.36)	(2.73)	(.17)	(.20)	(.24)
		1.0	1.0	.98	1.0	.99	.96	.98	.98	.98	.97	.97	.97	.97	.98	.97
C 4:		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)	(76.7)	(82.6)	(84)	(70.3)	(76.9)	(78)
GARCH(1,1)		1.0	1.0	.93	.99	.99	.94	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
C 5:		27.3	30.4	33.1	22.8	29.4	27	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	.98	1.0	1.0	1.0	1.0	.99	1.0	.99	.98	1.0	.99	.99	1.0	.99
C 6:		1.64	2.82	188.5	1.42	2.48	162.9	1.01	1.70	108.7	.99	1.64	104.2	1.03	1.68	105.67
Deb. des. mat.:		(.29)	(.50)	(32.3)	(.32)	(.51)	(34.4)	(.26)	(.36)	(25.6)	(.31)	(.43)	(27.7)	(.21)	(.33)	(24.5)
		1.0	1.0	1.0	1.0	1.0	1.0	.96	.98	.97	.94	.96	.96	.96	.96	.96
C 7:		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.:		(.59)	(.82)	(.82)	(.56)	(.78)	(.74)	(.32)	(.46)	(.45)	(.38)	(.57)	(.53)	(.37)	(.49)	(.47)
		1.0	1.0	1.0	1.0	1.0	.99	.98	.98	.98	.97	.98	.97	.99	.98	.99
		Models which do not satisfy the mediangale condition														
C 8:		1.59	1.71	1.46	1.63	1.47	1.05	.99	1.00	.94	1.17	.96	.86	1.23	.91	.81
$\rho_u = .5,\ \rho_x = 0$		(.30)	(.32)	(.30)	(.38)	(.31)	(.28)	(.25)	(.26)	(.24)	(.34)	(.26)	(.23)	(.36)	(.23)	(.21)
HOM		.99	1.0	1.0	.99	1.0	.99	.86	.98	.99	.90	.97	.97	.91	.96	.95
C 9:		1.25	1.46	1.56	1.23	1.64	.99	.68	1.12	.96	.79	1.23	.96	.94	1.11	1.01
$\rho_u = \rho_x = .5$		(.31)	(.40)	(.40)	(.41)	(.51)	(.35)	(.17)	(.33)	(.24)	(.24)	(.42)	(.26)	(.33)	(.55)	(.36)
HET		1.0	.99	1.0	.98	.97	.94	.93	.88	.98	.95	.89	.98	.97	.83	.97
C 10:		2.46	2.42	2.69	3.00	2.00	2.41	1.52	1.41	1.51	2.46	1.56	1.53	2.89	1.21	1.27
$\rho_u = .9,\ \rho_x = 0$		(.84)	(.82)	(.95)	(1.06)	(.68)	(.96)	(.57)	(.56)	(.61)	(1.00)	(.60)	(.63)	(1.46)	(.47)	(.61)
HOM		.68	.99	1.0	.74	1.0	.99	.47	.95	.98	.66	.97	.98	.71	.87	.91

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 4 last months of 1929 (90 observations), to investigate behaviors of the different methods in small samples. Due to the financial crisis, one may expect data to involve heavy heteroskedasticity during this period. Let us remind the Wall Street krach occurred between October 24 (*Black Thursday*) and October 29 (*Black Tuesday*). Hence, the second subsample corresponds to September, October with the krach period, and November and December with the early beginning of the Great Depression. Heteroskedasticity tests reject homoskedasticity for both subsamples.[16]

In Table 6, 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 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) either some bias is suspected (see OLS results on the whole period). Take the constant parameter, on the 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 toward 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 heteroskedasticity, least squares methods should be avoided.

Sign-based tests seem really adapted for small samples settings. Let us examine the third column of Table 6. 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 setup. This suggests the presence of autocorrelation in the disturbance process. In such a circumstance,

[16]1929: White: 24.2, p-values: .000 ; BP: 126, p-values: .000; Sept-Oct-Nov-Dec 1929: White: 11.08, p-values: .004; BP: 1.76, p-values: .18.

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.

8.2. β-convergence across U.S. States

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

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

$i = 1, \ldots, 48,\ T = 8,\ 10$ or $20,\ \ t = 1900,\ 1920,\ 1930,\ 1940, 1950,\ 1960,\ 1970,\ 1980,$ 1988. Their *basic equation* does not include any other variables but they also consider a specification with regional dummies (*Eq. with reg. dum.*). The *basic equation* assumes that the 48 States share a common per capita level of personal income at steady state while the second specification allows for regional differences in steady state levels. Their regressions involve 48 observations and are run for each 20-year or 10-year period between 1880 and 1988. They tended to accept a positive β and concluded on a convergence between levels of per capita personal income across U.S. States. However, both the NLLS method and the Wald-type tests they performed are only asymptotically justified and can be unreliable for only 48 observations. This unreliability is strengthened here because the data suffer from heteroskedasticity, departure from normality, presence of outliers and observations with possibly high influence. Indeed, residual analysis show that departures from a normal standard case are present in most periods (see Table 7).[17] Only, the outstanding growth period of 1960-1970 does not seem to show potential data problems. Similar results hold for the equation with regional dummies. This survey highly reduces the validity of least squares methods and suggests the need of a test, valid in finite samples and robust to heteroskedasticity of unknown form.

Hence, we propose to perform finite-sample based sign tests to see whether the conclusion of β-convergence still holds. We consider the linear equation:

$$(1/T)\ln(y_{i,t}/y_{i,t-T}) = a + \gamma[\ln(y_{i,t-T})] + x_i'\delta + \epsilon_i^{t,T} \tag{8.2}$$

where x_i contains regional dummies when included, and compute projection-based CI for γ, a, and for $\beta = -(1/T)\ln(\gamma T + 1)$ as a bijective transformation of γ, in both specifications. We compare projection-based valid 95%-confidence intervals for β based on the sign-based statistic SF with Barro and Sala-i-Martin nonlinear least squares asymptotic

[17] Omitted variables, misspecification of the model can also lead to similar conclusions, we do not consider those problems here, which leads to entirely rethink the growth theory and the model.

95%-confidence intervals (Table 8).

The results we find for the basic regression are close to those of Barro and Sala-i-Martin (1991). We fail to reject $\beta = 0$ at 5%-level, for the 1880-1900, 1920-1930, 1980-1988 periods, whereas Barro and Sala-i-Martin (1991) fail to reject $\beta = 0$ at 5% (asymptotic)-level for the 1920-1930 and 1980-1988 periods. Our results differ only for the 1880-1900 period. That may be due to the strong heteroskedasticity and departure from normality affecting least squares methods as we show in Table 7. When regional dummies are included, we fail to reject $\beta = 0$ at 5%-level 7 times over 9 whereas Barro and Sala-i-Martin (1991) fail to reject 3 times over 9. Finally, a positive β convergence seems to pass both NLLS-based asymptotic tests and finite sample-based robust sign tests with the basic specification, yielding to a strong argument in favor of the theory. However, that is no longer true for the specification with regional dummies, which reduces the idea of a strictly positive β convergence with possibly different regional steady state levels. This also may in part be due to the conservativeness of the projection-based method but there is no evidence that smaller exact confidence intervals can be constructed.

9. 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 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 heteroskedasticity is suspected in the data and the number of available observations is small. Finally we have presented two practical examples: we test the presence of a drift on the S&P price index, for the whole period 1928-1987 and for shorter subsamples. And, we reinvestigate whether a β-convergence between levels of per capita personal income across U.S. States occurred between 1880 and 1988.

Appendix

A. Proofs

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

Proof of Proposition 3.2. Consider the vector $[s(u_1), s(u_2), \ldots, s(u_n)]' \equiv (s_1, s_2, \ldots, s_n)'$. From Assumption 2.1, we derive the two following equalities:

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

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

$$l(s_1, s_2, \ldots, s_n|X) = \prod_{t=1}^{n} l(s_t|s_{t-1}, \ldots, s_1, X)$$

$$= \prod_{t=1}^{n} \mathsf{P}[u_t > 0|u_{t-1}, \ldots, u_1, X]^{(1-s_t)/2}\{1 - \mathsf{P}[u_t > 0|u_{t-1}, \ldots, u_1, X]\}^{(1+s_t)/2}$$

$$= \prod_{t=1}^{n} (1/2)^{(1-s_t)/2}[1 - (1/2)]]^{(1+s_t)/2} = (1/2)^n.$$

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

Proof of Proposition 3.3. Consider model (2.1) with $\{u_t : t = 1, 2, \ldots\}$ satisfying a weak mediangale conditional on X. Let show that $\tilde{s}(u_1), \tilde{s}(u_2), \ldots, \tilde{s}(u_n)$ can have the same role in Proposition **3.2** as $s(u_1), s(u_2), \ldots, s(u_n)$ under Assumption 2.1. The randomized signs are defined by $\tilde{s}(u_t, V_t) = s(u_t) + [1 - s(u_t)^2]s(V_t - .5),$ hence

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

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:

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

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

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

In a similar way,

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

The rest is similar to the proof of Proposition **3.2**. □

Proof of Proposition 4.1. 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

$$\begin{aligned}
\mathsf{P}_\beta[S(y) = s|X] &= \prod_{i=1}^{n} [\mathsf{P}_\beta(y_i > 0|X)]^{(1+s_i)/2} [\mathsf{P}_\beta(y_i < 0|X)]^{(1-s_i)/2} && \text{(A.5)} \\
&= \prod_{i=1}^{n} [1 - F_i(-x_i\beta|X)]^{(1+s_i)/2} [F_i(-x_i\beta|X)]^{(1-s_i)/2}. && \text{(A.6)}
\end{aligned}$$

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

$$\begin{aligned}
\mathsf{P}_\beta[S(y) = s|X] &= \frac{1}{2^n} \prod_{i=1}^{n} [1 + 2f_i(0|X)x_i s_i \beta + o(\beta)] && \text{(A.7)} \\
&= \frac{1}{2^n} \left[1 + 2\sum_{i=1}^{n} f_i(0|X)x_i s_i \beta + o(\beta) \right]. && \text{(A.8)}
\end{aligned}$$

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

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

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 power function around zero is determined by the quadratic term of its Taylor expansion:

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

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, Simonova, and Tiurin (p. 41, 1997), Dubrovin, Fomenko, and Novikov (ch. 2, pp. 76-86, 1984) or Gray (ch. 21, pp. 373-380,1998). Taking the trace in expression (A.11), we find (after some computations) that

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

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 heteroskedastic 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 6.1. 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 6.4, $V_n^{-1/2}$ 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 6.1 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 6.2 and 6.3 imply

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

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

Note also that

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

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

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

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

Proof of Corollary 6.2. 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, $V_n = \mathrm{Var}\left[\frac{1}{\sqrt{n}} s \otimes X\right] = \frac{1}{n}\sum_{t=1}^{n} \mathsf{E}(x_t s_t s_t' x_t') = \frac{1}{n}\sum_{t=1}^{n} \mathsf{E}(x_t x_t') = \frac{1}{n}\mathsf{E}(X'X)$, and $X'X/n$ is a consistent estimate of $\mathsf{E}(X'X/n)$. Theorem **6.1** yields $SF(\beta_0) \to \chi_2(p)$. □

In order to prove Theorem **6.3**, we will use the following lemma on the uniform convergence of distribution functions [see Chow and Teicher (1988, sec. 8.2, p. 265)].

Lemma A.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 6.3. $G(-\infty) = \tilde{G}_n(-\infty) = 0$, $G(+\infty) = \tilde{G}_n(+\infty) = 1$, and $\tilde{G}_n\big(x|X_n(\omega)\big) \to G(x)$ $a.e.$. By Lemma **A.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

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

hence

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

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 $\alpha \in Im\big(G_n(\mathbb{R}^+)\big)$, $i.e.$ for each point of the image set, we have

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

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

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

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

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

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

$$\mathsf{P}\big[G_n\big(c_n S_n(\beta_0)\big) \leq \alpha\big] = \mathsf{P}\big[G_n\big(c_n S_n(\beta_0)\big) \leq \alpha_1\big] = \alpha_1 \leq \alpha. \tag{A.22}$$

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

$$\mathsf{P}\big[\tilde{G}_n\big(c_n S_n(\beta_0)\big) \leq \alpha\big] = \mathsf{P}\big[G_n\big(c_n S_n(\beta_0)\big) \leq \alpha\big] + o_p(1) \leq \alpha + o_p(1).$$

□

Proof of Theorem 6.4. 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 $(c_n S_n^{(0)}, \ldots, c_n S_n^{(N)})$ is exchangeable. Indeed, in that case, the distribution of ranks is fully specified and yields the validity of empirical $p-$value [see Dufour (2006)]. In our case, it is clear that $(c_n S_n^{(0)}, \ldots, c_n S_n^{(N)})$ 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 $(c_n S_n^{(0)}, \ldots, c_n S_n^{(N)})$ is asymptotically exchangeable, we show that for any permutation $\pi : [1, N] \to [1, N]$,

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

First, let rewrite

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

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

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

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

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

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

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

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

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

Hence, we can apply an asymptotic version of Proposition 2.2.2 in Dufour (2006) that validates Monte Carlo testing for general possibly noncontinuous 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, $\lim_{n\to\infty} \tilde{p}_n^N(c_n S_n^{(0)}) \leq \alpha$. $\square$

B. Detailed analysis of Barro and Sala-i-Martin data set

This appendix contains additional results for the Barro and Sala-i-Martin application. Table 9 contains results of heteroskedasticity tests. Complementary sign-based inference results for the model parameters are reported in Table 10.

Table 6. S&P price index: 95 % confidence intervals

Constant parameter (a) Methods	Whole sample *(16120 obs)*	Subsamples *1929 (291 obs)*	*1929 (90 obs)*
Sign			
SF statistics	[-.007, .105]	[-.226, .522]	[-1.464, .491]
SHAC statistics	[-.007, .106]	[-.135, .443]	[-.943, .362]
LAD (estimate)	*(.062)*	*(.163)*	*(-.091)*
with OS cov. matrix est.	[.033, .092]	[-.144, .470]	[-1.015, .832]
with DMB cov. matrix est.	[.007, .117]	[-.139, .464]	[-1.004, .822]
with MBB cov. matrix est. (b=3)	[.008, .116]	[-.130, .456]	[-1.223, 1.040]
with kernel cov. matrix est. (Bn=10)	[-.019, .143]	[-.454, -.780]	[-1.265, 1.083]
OLS	*(-.005)*	*(.224)*	*(-.522)*
with iid cov. matrix est.	[-.041, .031]	[-.276, .724]	[-2.006, .962]
with DMB cov. matrix est.	[-.054, .045]	[-.142, .543]	[-1.335, .290]
with MBB cov. matrix est. (b=3)	[-.056, .046]	[-.140, .588]	[-1.730, .685]
Drift parameter (b) Methods	$\times 10^{-5}$	$\times 10^{-2}$	$\times 10^{-1}$
Sign			
SF statistics	[-.676, .486]	[-.342, .344]	[-.240, .305]
SHAC statistics	[-.699 , .510]	[-.260, .268]	[-.204, .224]
LAD	*(.184)*	*(.000)*	*(-.044)*
with OS cov. matrix est.	[-.504 , .320]	[-.182, .182]	[-.220, .133]
with DMB cov. matrix est.	[-.688 , .320]	[-.256, .255]	[-.281, .194]
with MBB cov. matrix est. (b=3)	[-.681 , .313]	[-.236, .236]	[-.316, .229]
with kernel cov. matrix est.	[-.671, -.104]	[-.392, .391]	[-.303, .215]
OLS	*(.266)*	*(-.183)*	*(.010)*
with iid cov. matrix est.	[-.119 , .651]	[-.480, .113]	[-.273, .293]
with DMB cov. matrix est.	[-.213 , .745]	[-.544, .177]	[-.148, .169]
with MBB cov. matrix est. (b=3)	[-.228 , .761]	[-.523, .156]	[-.250, .270]

Table 7. Regressions for personal income across U.S. States: regression diagnostics

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

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

Table 8. Regressions for personal income across U.S. States: 95% -confidence intervals

Period		*Basic equation*		*Eq. with reg. dum.*	
	β	SIGN (SF)	*NLLS**	SIGN (SF)	*NLLS**
1880-1900:	[*95%CI*] (β^{NLLS})	[-.0010, .0208]	*[.0058, .0532] (.0101)*	[-.0033, .0251]	*[.0146, .0302] (.0224)*
1900-1920:		[.0092, .0313]	*[.0155, .0281] (.0218)*	[-.0081, .0558]	*[.0086, .0332] (.0209)*
1920-1930:		[-.0301, .0018]	*[-.0249, -.0049] (-.0149)*	[-.0460, .0460]	*[-.0267, .0023] (-.0122)*
1930-1940:		[.0043, .0234]	*[.0082, .0200] (.0141)*	[-.0187, .0377]	*[.0027, .0227] (.0127)*
1940-1950:		[.0291, .0602]	*[.0372, .0490] (.0431)*	[.0082, .0620]	*[.0314, .0432] (.0373)*
1950-1960:		[.0084, .0352]	*[.0121, .0259] (.0190)*	[.0007, .0506]	*[.0100, .0304] (.0202)*
1960-1970:		[.0099, .0377]	*[.0170, .0322] (.0246)*	[-.0112, .0431]	*[.0047, .0215] (.0131)*
1970-1980:		[.0021, .0346]	*[.0076, .0320] (.0198)*	[-.0227, .0721]	*[-.0016, .0254] (.0119)*
1980-1988:		[-.0552, .0503]	*[-.0315, .0195] (-.0060)*	[-.0467, .0754]	*[-.0273, .0173] (-.0050)*

* Barro and Sala-i-Martin (1991) NLLS results are reported in those two columns.

Table 9. Regressions for personal income across U.S. States, 1880-1988: tests for heteroskedasticity.

Period	*Basic equation*		*Eq. with reg. dum.*	
p-values	White test	Breusch-Pagan test	White test	Breusch-Pagan test
1880-1900	.018	.652	.249	.830
1900-1920	.023	.043	.069	.050
1920-1930	.723	.398	.435	.557
1930-1940	.673	.633	.537	.601
1940-1950	.243	.943	.513	.272
1950-1960	.595	.223	.740	.221
1960-1970	.205	.247	.236	.441
1970-1980	.641	.675	.777	.264
1980-1988	.058	.022	.080	.226

Table 10. Regressions for personal income across U.S. States, 1880-1988: complementary results.

Period	*Basic equation*	*Eq. with reg. dum.*
Variable: constant (a)	*95% projection-based CI(a)*	
1880-1900	[-.0147, -.0020]	[.0206, .0005]
1900-1920	[-.0205, -.0084]	[-.0431, .0095]
1920-1930	[-.0018, .0328]	[-.0351, .0589]
1930-1940	[-.0232, -.0042]	[-.0443, .0221]
1940-1950	[-.0452, -.0258]	[-.0517, -.0070]
1950-1960	[-.0297, -.0080]	[-.0435, .0043]
1960-1970	[-.0314, .0088]	[-.0345, .0119]
1970-1980	[-.0296, -.0020]	[-.0478, .0288]
1980-1988	[-.0414, .0695]	[-.0563, .0566]
Variable: $\ln(y)$ (γ)	*95% projection-based CI(a)*	
1880-1900	[-.0170, .0010]	[-.0197, .0034]
1900-1920	[-.0233, -.0084]	[-.0336, .0088]
1920-1930	[-.0018, .0351]	[-.0369, .0584]
1930-1940	[-.0209, -.0042]	[-.0314, .0206]
1940-1950	[-.0452, -.0253]	[-.0462, .0079]
1950-1960	[-.0297, -.0080]	[-.0397, -.0007]
1960-1970	[-.0314, -.0094]	[-.0350, .0119]
1970-1980	[-.0292, -.0020]	[-.0514, .0255]
1980-1988	[-.0414, .0695]	[-.0566, .0566]

C. Compared inference methods in simulations

Two *sign-based statistics* are studied ins Section 7: one adapted for the mediangale process,

$$SF(\beta_0) = D_S(\beta_0,\, (X'X)^{-1}) = s(y - X\beta_0)'X(X'X)^{-1}X's(y - X\beta_0) \tag{C.1}$$

and one corrected for serial dependence,

$$SHAC = D_S(\beta_0,\, \hat{J}_n^{-1}) = s(y - X\beta_0)'X\hat{J}_n^{-1}X's(y - X\beta_0) \tag{C.2}$$

where

$$\hat{J}_n = \frac{n}{n-p} \sum_{j=-n+1}^{n-1} k\left(\frac{j}{B_n}\right) \hat{\Gamma}_n(j), \tag{C.3}$$

with

$$\hat{\Gamma}_n(j) = \begin{cases} \frac{1}{n}\sum_{t=j+1}^{n} V_t(\beta_0)V'_{t-j}(\beta_0) & \text{for } j \geq 0 \\ \frac{1}{n}\sum_{t=-j+1}^{n} V_{t+j}(\beta_0)V'_t(\beta_0) & \text{for } j < 0, \end{cases} \tag{C.4}$$

and $V_t(\beta_0) = s(y_t - x'_t\beta_0) \times x_t,\ t = 1, \ldots, n$ and $k(\cdot)$ is a real-valued kernel, here Bartlett kernel is used with an automatically adjusted bandwidth parameter B_n [Andrews (1991)].

Sign-based tests are compared to LR and Wald-type tests based on OLS and LAD estimators with different covariance matrix estimators. Wald-type statistics for testing $H_0(\beta_0) : \beta = \beta_0$ are of the form $n(\hat{\beta} - \beta_0)\hat{D}_n^{-1}(\hat{\beta} - \beta_0)$ where $\hat{D}_n$ is an estimate of the asymptotic covariance matrix for $\hat{\beta}$.

The OLS estimator is computed in GAUSS: $\hat{\beta}_{OLS} = (X'X)^{-1}X'y$. Both *classic i.i.d.* and *White covariance matrix estimators* are considered. WH asymptotic covariance matrix estimator is corrected for heteroskedasticity but not for linear dependence:

$$\hat{D}^{WH}(\hat{\beta}_{OLS}) = \left(\frac{1}{T}\sum x_t x'_t\right)^{-1} \left(\frac{1}{T(T-k)}\sum \hat{u}_t^2 x_t x'_t\right) \left(\frac{1}{T}\sum x_t x'_t\right)^{-1}.$$

The LAD estimator is computed in GAUSS by the qreg procedure, which uses a minimization by interior point method: $\hat{\beta}_{LAD} = \arg\min \sum_{t=1}^{n} |y_t - x'_t\beta|$. The following LAD covariance matrix estimators are considered:
The order statistic estimator (OS) [see Chamberlain (1994), Buchinsky (1995, 1998)], which is valid for $i.i.d$ observations, is used as a benchmark. For $i.i.d$ observations, the LAD covariance matrix reduces to

$$D(\hat{\beta}_{LAD}) = \frac{1}{4f_u^2(0)}(E[xx'])^{-1} = \sigma_{LAD}^2(E[xx'])^{-1},$$

where f_u stands for the density of u_t. An estimate for σ_{LAD} can be constructed from a confidence interval for the sample median, $i.e.$, the $n/2$-th order statistic. let $y_1, y_2, \ldots, y_n$ be independent random observations with distribution function $F_y(.)$ and $y_{(j)}, y_{(k)}$, the $j-$th and the $k-$th order statistics of $y_1, y_2, \ldots, y_n$. Note that $\mathrm{P}[y_{(j)} \leq \xi_{1/2}] = \sum_{i=j}^{n} C_n^i (1/2)^n$, which entails

$$\mathrm{P}[y_{(j)} \leq \xi_{1/2} \leq y_{(k)}] = \mathrm{P}[y_{(j)} \leq \xi_{1/2}] - \mathrm{P}[y_{(k)} < \xi_{1/2}] = \sum_{i=j}^{k-1} C_n^i (1/2)^n.$$

A symmetric confidence interval with level $1-\alpha$ can be constructed as follows. Let $j = int(n/2-l)$, $k = int(n/2+l)$ and $X \sim \mathcal{B}(n, 1/2)$, with $E[X] = n/2$ and $var(X) = n/4$. Then,

$$\begin{aligned} \mathbf{P}[Y_{int(n/2-l)} \leq \xi_{1/2} \leq Y_{int(n/2+l)}] &= \mathbf{P}[int(n/2) - l \leq X \leq int(n/2) + l] \\ &= \mathbf{P}\left[\frac{X - n/2}{\sqrt{n/4}} \leq \frac{l}{\sqrt{n/4}}\right]. \end{aligned}$$

A central limit theorem, $\frac{X-n/2}{\sqrt{n/4}} \to \mathcal{N}(0,1)$ entails that $l = Z_{1-\alpha/2}\sqrt{n/4}$ where $Z_{1-\alpha/2}$ is the $1-\alpha/2$th quantile of a standard normal distribution. Approaching the width of the exact confidence interval by that of asymptotic confidence interval gives $\hat{\sigma}^2_{LAD} = \frac{n(Y_{int(n/2+l)} - Y_{int(n/2-l)})^2}{4Z^2_{1-\alpha/2}}$. Finally, $D(\hat{\beta}_{LAD})$ can be estimated by,

$$\hat{D}^{OS}(\hat{\beta}_{LAD}) = \hat{\sigma}^2_{LAD} \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1}.$$

Design matrix bootstrap centering around the sample LAD estimate (DMB) is also considered [see Buchinsky (1995, 1998)]. Let $(y_i^*, x_i^*),\ i = 1, \ldots, m$ be a randomly drawn sample from the empirical distribution function F_{nxy}. Let $\hat{\beta}^*_{LAD}$ be the bootstrap estimate obtained from a LAD regression of y^* on X^*. This process is carried out B times and yields B bootstrap estimates, $\hat{\beta}^*_{LAD1},\ \hat{\beta}^*_{LAD2}, \ldots,\ \hat{\beta}^*_{LADB}$. The design matrix bootstrap asymptotic covariance matrix estimator is given by,

$$\hat{D}^{DMB} = \frac{m}{n}\left\{\frac{n}{B}\sum_{j=1}^{B}(\hat{\beta}^*_{LADj} - \hat{\beta}_{LAD})(\hat{\beta}^*_{LADj} - \hat{\beta}_{LAD})'\right\}. \quad \text{(C.5)}$$

The moving block bootstrap centering around the sample estimate (MBB) was proposed by Fitzenberger (1997*b*). Basically, blocks of fixed size b are bootstrapped instead of individual observations. $q = T - b + 1$ blocks of observations of size b,

$B_i = ((y_i, x_i), \ldots, (y_{i+b}, x_{i+b}))$ are defined. m blocks, drawn from the initial sample, constitute a bootstrapped sample Z_j of size $m \times b$. From each $Z_j,\ j = 1, \ldots, B$, a LAD regression is performed yielding the estimate $\hat{\beta}^{*j}_{LAD}$. The MBB estimator of the LAD asymptotic covariance matrix can then be approached thanks to the bootstrap paradigm, by

$$\hat{D}^{MBB}(\hat{\beta}_{LAD}) = \frac{mb}{B} \left\{ \sum_{j=1}^{B} (\hat{\beta}^{*}_{LADj} - \hat{\beta}_{LAD})(\hat{\beta}^{*}_{LADj} - \hat{\beta}_{LAD})' \right\}. \tag{C.6}$$

Both for OLS and LAD estimators *Bartlett kernel covariance matrix estimators with automatic bandwidth parameter* (BT) are also considered [see Parzen (1957), Newey and West (1987), Andrews (1991)] with a methodology similar to the one presented previously for deriving the $SHAC$-sign statistic.

Finally, the LR *statistic* [see Koenker and Bassett (1982)] has the following form:

$$4\hat{f}_u(0) \left[\sum |y_i - x_i'\beta_0| - \sum |y_i - x_i'\hat{\beta}_{LAD}| \right] \tag{C.7}$$

where an OS estimate is used for $\hat{f}_u(0)$.

References

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

Amemiya, T. (1982), 'Two stage least absolute deviations estimators', *Econometrica* **50**(3), 689–712.

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

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

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

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

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

Buchinsky, M. (1994), 'Changes in the U.S. wage structure 1963-1987: Application of quantile regression', *Econometrica* **62**, 405–458.

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

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

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

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

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

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

Chamberlain, G. (1994), Quantile regression, censoring and the structure of wages, *in* C. A. Sims, ed., 'Advances in Econometrics', Econometric Society Monographs, Elsevier, New York, pp. 171–209.

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

Chernozhukov, V., Hansen, C. and Jansson, M. (2006), Finite sample inference for quantile regression models, Technical Report 06-03, Department of Economics, M.I.T., Cambridge, MA.

Chow, Y. S. and Teicher, H. (1988), *Probability Theory. Independence, Interchangeability, Martingales. Second Edition*, Springer-Verlag, New York.

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

Coudin, E. and Dufour, J.-M. (2007), Generalized confidence distributions and robust sign-based estimators in median regressions under heteroskedasticity and nonlinear dependence of unknown form, Discussion paper, CIREQ, Université de Montréal and CREST-INSEE.

David, H. A. (1981), *Order Statistics*, second edn, John Wiley & Sons, New York.

De Angelis, D., Hall, P. and Young, G. A. (1993), 'Analytical and bootstrap approximations to estimator distributions in $L1$ regression', *Journal of the American Statistical Association* **88**(424), 1310–1316.

Dielman, T. and Pfaffenberger, R. (1988*a*), 'Bootstrapping in least absolute value regression: An application to hypothesis testing', *Communications in Statistics-Simulation and Computation* **17**(3), 843–856.

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

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

Dubrovin, B., Fomenko, A. and Novikov, S. (1984), *Modern-Geometry- Methods and Applications*, Graduate texts in Mathematics, Springer-Verlag, New-York.

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

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

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

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

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

Dufour, J.-M. and Kiviet, J. F. (1998), 'Exact inference methods for first-order autoregressive distributed lag models', *Econometrica* **66**, 79–104.

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

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

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

Engle, R. and Manganelli, S. (2000), CAViaR: Conditional autoregressive value at risk by regression quantiles, Technical report, Paper presented at Econometric Society World Congress 2000.

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

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

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

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

Goffe, W. L., Ferrier, G. D. and Rogers, J. (1994), 'Global optimization of statistical functions with simulated annealing', *Journal of Econometrics* **60**, 65–99.

Gouriéroux, C. and Monfort, A. (1995), *Statistics and Econometric Models, Volumes One and Two*, Cambridge University Press, Cambridge, U.K.

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

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

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

Hallin, M. and Puri, M. L. (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', Vol. 45 of *The IMA Volumes in Mathematics and its Applications*, Springer-Verlag, New York, pp. 111–153.

Hallin, M., Vermandele, C. and Werker, B. (2006), 'Linear and nonserial sign-and-rank statistics: Asymptotic representation and asymptotic normality', *The Annals of Statistics* **34**, 254–289.

Hallin, M., Vermandele, C. and Werker, B. (2008), 'Semiparametrically efficient inference based on signs and ranks for median-restricted models', *Journal of the Royal Statistical Society, Series B* **70**(2), 389–412.

Hallin, M. and Werker, B. (2003), 'Semiparametric efficiency, distribution-freeness, and invariance', *Bernoulli* **9**(1), 137–165.

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

Horowitz, J. L. (1998), 'Bootstrap methods for median regression models', *Econometrica* **66**, 1327–1351.

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

Kim, T. H. and White, H. (2002), Estimation, inference, and specification testing for possibly misspecified quantile regression, Technical report, Department of Economics, Univrsity of California at San Diego, San Diego, California.

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

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

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

Koenker, R. and Hallock, K. (2001), 'Quantile regression', *Journal of Economic Perspectives* **15**(4), 141–156.

Komunjer, I. (2005), 'Quasi-maximum likelihood estimation for conditional quantiles', *Journal of Econometrics* **128**(1), 137–164.

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

Mankiw, N. G. and Shapiro, M. D. (1986), 'Do we reject too often? Small sample properties of rational expectations models', *Economics Letters* **20**, 243–247.

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

Parzen, E. (1957), 'On consistent estimates of the spectrum of a stationary times series', *Annals of Mathematical Statistics* **28**(2), 329–348.

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

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

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

Powell, J. L. (1994), Estimation of semiparametric models, *in* R. F. Engle and D. L. McFadden, eds, 'Handbook of Econometrics, Volume 4', North-Holland, Amsterdam, chapter 41, pp. 2443–2521.

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

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1996), *Numerical Recipes in Fortran 90*, second edn, Cambridge University Press, Cambridge, U.K.

Scheffé, H. and Tukey, J. W. (1945), 'Non-parametric estimation. I. validation of order statistics', *Annals of Mathematical Statistics* **16**(2), 187–192.

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**(3), 122–128.

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

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

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

White, H. (2001), *Asymptotic Theory for Econometricians*, revised edn, Academic Press, San Diego, Florida.

Wright, J. H. (2000), 'Alternative variance-ratio tests using ranks and signs', *Journal of Business and Economic Statistics* **18**(1), 1–9.

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