Inférence exacte et non paramétrique dans les modèles de régression et les modèles structurels en présence d'hétéroscédasticité de forme arbitraire

Par

Elise Coudin

Thèse de doctorat effectuée en cotutelle

au

Département de Sciences économiques

Faculté des arts et des sciences

et à

l'Ecole des Hautes Etudes en Sciences Sociales

Thèse présentée à la Faculté des études supérieures

en vue de l'obtention du grade de Philosophiae Doctor (Ph.D.)

en sciences économiques

et à

l'Ecole des Hautes Etudes en Sciences Sociales

en vue de l'obtention du grade de Docteur

Juin 2007

Page d'identification du jury

Université de Montréal

Faculté des études supérieures

et

Ecole des Hautes Etudes en Sciences Sociales

Cette thèse intitulée :

Inférence exacte et non-paramétrique dans les modèles de régression et les modèles structurels en présence d'hétéroscédasticité de forme arbitraire

Présentée et soutenue à l'Université de Montréal par :

Elise Coudin

a été évaluée par un jury composé des personnes suivantes :

Président-rapporteur et membre du jury	William J. McCausland
Directeur de recherche (Université de Montréal)	Jean-Marie Dufour
Directeur de recherche (EHESS)	Thierry Magnac
Examinateurs externes	Marc Hallin
	Rep: Benoit Perron
Membres du jury	Marine Carrasco
Représentant du doyen de la FES	Christian Léger

Sommaire

L'objet de cette thèse est de développer un système d'inférence exacte en échantillon fini dans des modèles de régression et des modèles structurels sans imposer d'hypothèse paramétrique sur la distribution des erreurs.

Dans le premier essai, nous étudions la construction de tests et de régions de confiance dans une régression linéaire sur la médiane. Le modèle que nous considérons n'impose pas de restriction paramétrique sur la distribution des erreurs. Celles-ci peuvent être non gaussiennes, hétéroscédastiques ou bien présenter une dépendance sérielle de forme arbitraire. Habituellement, l'analyse de ce type de modèle a recours à des approximations asymptotiques normales, lesquelles peuvent être trompeuses en échantillon fini. Nous introduisons une propriété analogue à la différence de martingale pour la médiane, la « mediangale », et remarquons que les signes d'une suite de « mediangale » sont indépendants entre eux et suivent une distribution connue et simulable. Nous utilisons alors 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. Grâce à la méthode des tests de Monte Carlo et à celle des projections, nous construisons tour à tour des tests et des régions de confiance simultanés puis des tests et des régions de confiance pour n'importe quelle tranformation du paramètre. Nous fournissons ensuite une théorie asymptotique sous des hypothèses plus faibles que la « mediangale ». Les études par simulation montrent que la méthode proposée est plus performante que les méthodes asymptotiques habituelles lorsque le processus est très hétérogène ou lorsque la taille de l'échantillon est petite. Enfin, deux exemples d'application sont étudiés. Dans le premier, nous testons la présence d'une tendance sur des données financières. Le deuxième s'appuie sur des données régionales, nécéssairement peu nombreuses, pour tester la théorie macroéconomique de β convergence entre les niveaux de production des états américains.

Dans le deuxième essai, nous introduisons un estimateur et des outils d'inférence valides en échantillon fini moins communément utilisés. Nous étudions, tout d'abord, la fonction p-value qui associe un *degré de confiance* à chaque valeur testée du paramètre étant donnée la réalisation de l'échantillon. Celle-ci est reliée à la notion de distribution de confiance et aux distributions fiducielles de Fisher [Fisher (1930)]. Ces outils fournissent

un équivalent fréquentiste aux distributions bayésiennes *a posteriori*. Nous calculons des fonctions p-value simulées à partir de tests de Monte Carlo simultanés, puis des versions projetées pour chaque composante individuelle du paramètre. Nous suivons ensuite le principe d'inversion de test de Hodges et Lehmann [Hodges et Lehmann (1963)] et proposons d'utiliser comme estimateur, la valeur du paramètre associée au plus haut degré de confiance (à la plus forte p-value). L'estimateur de signe qui en découle est sans biais pour la médiane quand les erreurs sont symétriques, et il partage les propriétés d'équivariance de l'estimateur des moindres valeurs absolues (« Least Absolute Deviations, LAD »). Il est aussi convergent et asymptotiquement normal sous des conditions plus faibles que l'estimateur LAD. En échantillon fini, les simulations suggèrent qu'il est plus performant en termes de biais et d'erreur quadratique moyenne pour des processus très hétérogènes. Ces outils permettent de compléter l'analyse des deux exemples empiriques étudiés précédemment.

Dans le troisième essai, nous développons une approche inférencielle exacte en échantillon fini pour des modèles structurels non-linéaires. Nous proposons une version de la propriété de pivotalité des signes adaptée à un modèle instrumental. Les tests exacts qui en découlent ne dépendent pas du degré d'identification du paramètre. Ils sont en particulier valides en présence d'instruments faibles, pour des erreurs possiblement hétéroscédastiques et non gaussiennes. L'approche que nous proposons fait intervenir des régressions artificielles où l'on régresse les signes contraints sur des instruments auxiliaires dans l'esprit d'Anderson et Rubin [Anderson et Rubin (1949), Dufour (2003)]. Nous étudions de plus la question des instruments optimaux à inclure dans le modèle, ce qui permet de gagner de la puissance en cas de suridentification. Les simulations montrent que notre approche est plus performante que les méthodes usuelles (y compris celles qui sont robustes à la présence d'instruments faibles) lorsque les erreurs sont non gaussiennes, hétéroscédastiques et lorsque l'échantillon est petit. Cette méthode est utilisée sur les données de Angrist et Krueger (1991) pour analyser les rendements de l'éducation sur le salaire.

Mots clés : inférence exacte ; régression sur la médiane : régression quantile ; test de signe ; hétéroscédasticité ; non normalité ; dépendance ; test de Monte Carlo ; techniques de projection ; distribution de confiance ; endogénéité ; modèle structurel ; modèle non-linéaire ; instrument ; instrument faible ; convergence.

Summary

The objective of this thesis is to develop a whole system of exact inference in finite samples, for regression models and structural econometric models under very weak distributional assumptions on the error term.

In the first essay, we study the construction of finite-sample distribution-free tests and confidence sets for the parameters of a linear median regression when no parametric assumption is imposed on the noise distribution. The setup we consider allows for non-normality, heteroskedasticity and nonlinear serial dependence of unknown forms. Such semiparametric models are usually analyzed using asymptotically justified approximate methods, which can be arbitrarily unreliable in finite samples. We consider first the property of mediangale – the median-based analogue of a martingale difference – and show that the signs of mediangale sequences are distribution-free despite the presence of nonlinear dependence and heterogeneity of unknown form. We point out that a simultaneous inference approach in conjunction with sign transformations does provide statistics with the required pivotality features – in addition to usual robustness properties. Those sign-based statistics are exploited – with Monte Carlo tests and projection techniques – in order to produce valid inference in finite samples: simultaneous tests, confidence regions and then more general projection-based tests are constructed. An asymptotic theory which holds under even weaker assumptions is also provided. Simulations suggest the good performance of that method for a wide range of processes. Finally, two illustrative examples are presented. First, we test for the presence of a drift in financial series involving strong heteroskedasticity. Then, we exploit a cross-regional data set whose sample size is necessarily small, and test for β convergence between levels of per capita output across U.S. States.

The second essay presents additional finite-sample-based tools that can be used in conjunction with the sign-based inference system previously developed. First, we study the p-value function which measures the confidence one may have in a certain value of the parameter. It is related to the notion of confidence distribution and to Fisher fiducial distributions [Fisher (1930)]. Those notions provide a frequentist analogue to the Bayesian posterior distributions. We combine sign-based Monte Carlo tests of simultaneous hypotheses with projection techniques to construct simulated p-value functions and projected versions for the parameter individual components. Second, sign-based estimators that are the parameter values with the highest confidence (the highest p-value) are presented. These are obtained using the Hodges-Lehmann principle of test inversion [Hodges and Lehmann (1963)]. They are expected to present the same robustness properties than the test statistics from which they are derived and can directly be associated with the exact inference procedures described in the first essay. We also show they are median unbiased (under a symmetry assumption) and present equivariance features similar to the LAD estimator. Consistency and asymptotic normality are also provided under regularity conditions weaker than the ones required for the LAD estimator. In a simulation study of bias and root mean square error (RMSE), we find that sign-based estimators perform better than the LAD estimator in settings with sizable heteroskedasticity. Sign-based estimators and p-value functions are then used to complete the analysis of the two practical examples studied previously.

The third essay develops finite-sample distribution-free exact inference in nonlinear structural models. We propose an adapted version of the sign invariance that allows one to construct exact tests. We notice that the validity of those tests does not depend on identification assumptions nor on parametric approximations imposed on the errors. Sign-based tests equal the nominal size for any given sample size in presence of weak instruments, with non-normal and heteroskedastic errors. Basically, the sign-based approach relies on artificial regressions where the signs of the constrained residuals are regressed on some "auxiliary" instruments [Anderson and Rubin (1949), Dufour (2003)]. We also study the problem of building optimal instruments, which can lead to considerable gain of power in case of overidentification. Simulations show that sign-based methods overcome usual methods and methods robust to weak instruments in non-normal and heteroskedastic settings. A re-analysis of the returns to education based on Angrist and Krueger (1991) data is also provided.

Key words: exact inference; median regression; quantile regression; sign test; heteroskedasticity; non-normality; dependence; Monte Carlo test; projection techniques; confidence distribution; endogenenity; structural models; nonlinear models; instrument; weak instrument; consistency.

Table des matières

Liste des tableaux

Table des figures

Remerciements

Je tiens tout d'abord à remercier tous ceux que je risque d'oublier par la suite. A coup de petits riens, nombreux sont ceux qui ont contribué à ce travail.

Je remercie mes directeurs. D'abord pour avoir bien voulu tenter l'expérience de la cotutelle. Et surtout, pour m'avoir guidée dans le monde de l'économétrie. Je tiens à exprimer toute ma gratitude à Jean-Marie Dufour. Son soutien et sa générosité m'ont accompagné tout au long de cette thèse. C'est un honneur pour moi d'avoir pu travailler aux côtés de quelqu'un d'aussi brillant. Merci à Thierry Magnac pour ses conseils, son aide et sa confiance. Je regrette que les circonstances ne nous aient pas amenés à collaborer plus directement. Je retiendrai de leurs enseignements, entre autres, qu'un bon chercheur doit être armé d'une bonne dose de patience mais qu'il doit aussi être capable de se lancer.

Je remercie Marc Hallin et Frédéric Jouneau d'avoir accepté d'être examinateur externe et rapporteur de cette thèse. Je remercie Marine Carrasco, Silvia Gonçalves et Alain Trognon d'avoir accepté de composer mon jury, même au prix d'un voyage à Montréal.

Je remercie les laboratoires et institutions qui m'ont accueillie et financée : le CRDE-CIREQ et le département de sciences économiques de l'université de Montréal, puis, le laboratoire de microéconométrie appliquée du CREST et l'INSEE qui m'a aidée à continuer ma recherche après ma réussite au concours et soutenue dans ce projet. Je pense aux autres administrateurs de l'INSEE qui ont aussi entrepris une thèse. Je les encourage dans cette démarche comme ils l'ont fait pour moi.

Je remercie les participants aux séminaires et aux conférences auquels j'ai pris part pour leurs remarques constructives. Merci à Nour Meddahi pour ses conseils sur le deuxième chapitre ; à Hélène et Jan Johannes qui, en plein mois d'août, ont plongé dans une version préliminaire du troisième et à Charlotte Hardouin, pour ses conseils téléphoniques mathématiques. Merci à Pascale Valéry qui m'a fourni les données de l'application sur l'indice $S\&P$.

Je remercie Pascal Martinolli et ses prouesses bibliothèquesques. Sans lui, elle (en) serait loin cette thèse !

Je pense aux gens que j'ai rencontrés au cours de ces années. Enseignants, chercheurs et étudiants, avec lesquels j'ai sympathisé (la liste est longue...). Et, parmi eux tous mes co-bureaux ! Merci pour nos discussions, votre ambiance et vos sourires !

Je remercie Lucie, Lyne et Luc qui m'ont apporté un vrai chez-nous (acado-)québécois, c'est-à-dire rempli de chaleur.

Enfin, je remercie Etienne et ses multiples facettes qui vont de la hot-line informatique à l'intendance domestique, en passant par une intarissable source de « niaque ».

Introduction

Tout le monde croit aux erreurs normales, disait Henri Poincaré[1], les mathématiciens parce qu'ils s'imaginent que c'est un fait d'observation, et les observateurs que c'est un théorème mathématique.

Quand en économétrie, on relâche l'hypothèse de normalité des erreurs, c'est très souvent pour y revenir en ayant recours à des approximations asymptotiques. Ainsi, l'inférence « à la Wald » est couramment utilisée : on calcule un estimateur, puis son comportement asymptotique grâce à un théorème central limite, on en déduit ensuite des tests et des régions de confiance asymptotiques.

Pourtant, de nombreuses études empiriques ou par simulation soulignent les limites des approximations asymptotiques. Les exemples de tests qui présentent des distortions de niveau en échantillon fini sont nombreux [pour des exemples en séries temporelles, voir Dufour (1981), Campbell et Dufour (1995, 1997) et dans le contexte d'une régression sur la médiane, voir Buchinsky (1995), DeAngelis, Hall et Young (1993), Dielman et Pfaffenberger (1988a, 1988b)]. La normalité qu'elle soit imposée par le modèle paramétrique ou approchée en asymptotique ne vient donc pas toujours d'un fait d'observation.

Les limites des méthodes asymptotiques sont aussi bien connues dans la littérature statistique. On sait depuis Bahadur et Savage (1956) qu'il n'existe pas de procédure de test valide et puissante en échantillon fini pour tester une moyenne si on ne spécifie pas plus la forme de la distribution. La conséquence en est qu'à distance finie, un test basé sur la distribution asymptotique a une taille qui peut arbitrairement dévier de son niveau nominal. En d'autres termes, la moyenne n'est pas testable dans un modèle non paramétrique. Pour décrire une procédure de test qui soit valide en présence d'hétéroscédasticité de forme arbitraire, il est nécessaire de recourir à une mesure de localisation, comme la médiane.

Lehmann et Stein (1949) nous indiquent par ailleurs qu'il existe des procédures robustes à l'hétéroscédasticité de forme arbitraire : les procédures basées sur les signes. Ces deux résultats de la théorie des tests impliquent, entre autres, que les méthodes asymptotiques,

[1] Plus précisément, Henri Poincaré rapporte les dires d'un collègue dans la préface de son ouvrage *Thermodynamique*, 1908 : « un physicien éminent me disait un jour à propos de la loi des erreurs : tout le monde y croit fermement parce que les mathématiciens s'imaginent que c'est un fait d'observation, et les observateurs que c'est un théorème mathématique. »

en particulier celles qui s'appuient sur la moyenne, ne permettent pas de contrôler le niveau des tests en échantillon fini, et ce, même lorsque qu'elles sont dites « corrigées de l'hétéroscédasticité et de l'autocorrélation (HAC) ». Elles ne sont pas valides à distance finie. Utiliser la normalité asymptotique n'est donc pas toujours ce que conseillerait un théoricien.

Un autre épisode de l'histoire économétrique a renforcé la méfiance que peut inspirer l'inférence « à la Wald » : celui des *instruments faibles*. Lorsqu'un modèle structurel fait intervenir des variables explicatives endogènes, c'est-à-dire corrélées avec le terme d'erreur, on a habituellement recours à des méthodes instrumentales. Les instruments sont des variables auxiliaires exogènes, c'est-à-dire non corrélées avec le terme d'erreur, qui vont assurer l'identification des paramètres du modèle et permettre d'inférer sur leurs valeurs. Pour ce faire, ils doivent être pertinents, c'est-à-dire bien corrélés avec les variables explicatives endogènes. Lorsqu'ils ne le sont que *faiblement*, ils ne permettent pas de retrouver une bonne identification du paramètre (cas de non-identification ou de quasi-non-identification). En présence d'instruments faibles ou en l'absence d'identification, les statistiques de type Wald ont des comportements asymptotiques inhabituels et les tests asymptotiques qui en découlent ne sont pas valides.

La littérature sur les instruments faibles met fortement en garde contre les défauts des méthodes asymptotiques habituelles qui s'appuient sur une hypothèse d'identification et sur la normalité asymptotique des estimateurs. Elle rappelle aussi qu'il existe des statistiques pivotales robustes aux problèmes d'identification à partir desquelles on peut construire des tests valides. La première d'entre elles est la statistique d'Anderson et Rubin (AR) [Anderson et Rubin (1949)]. D'autres ont suivi [Kleibergen (2002, 2005, forthcoming), Moreira (2001, 2003), voir aussi Dufour et Jasiak (2001), Stock et Wright (2000), Dufour et Taamouti (2005), ...]. Cette littérature amène à réfléchir sur la mise en oeuvre de l'inférence. Elle remet l'accent sur l'importance des statistiques pivotales.

Partir des tests et d'une statistique pivotale pour en dériver un système d'inférence est classique en statistique. Cette démarche permet de plus de redécouvrir différentes notions moins communément utilisées en économétrie.

Encore faut-il que de tels pivots soient disponibles. C'est à cela que répond le résultat de Lehmann et Stein : en échantillon fini, dans un modèle avec hétéroscédasticité de forme arbitraire, une transformation par les signes peut aider à construire des pivots.

Dans cette thèse, nous proposons un système d'inférence exacte en échantillon fini pour des modèles de régression semi-paramétriques sur la médiane. A partir de statistiques pivotales basées sur les signes des résidus, nous construisons des tests de Monte Carlo [Dwass (1957), Barnard (1963), Dufour (2006)] qui exploitent la distribution *exacte* de ces statistiques. Le niveau de ces tests simultanés est contrôlé quelle que soit la taille de l'échantillon et ce, pour des formes arbitraires d'hétéroscédasticité et de dépendance non-linéaire. Nous construisons ensuite des régions de confiance simultanées en inversant ces tests, et des tests d'hypothèses plus générales grâce à des techniques de projection [Dufour et Kiviet (1998), Dufour et Jasiak (2001), Dufour et Taamouti (2005)]. Nous étudions ensuite d'autres outils d'inférence qui ont jusqu'à présent reçu moins d'écho dans la littérature économétrique : la fonction p-value et la distribution de confiance. L'estimateur constitue enfin la dernière brique de ce système d'inférence. Cette approche inférentielle commence donc par les tests et finit par l'estimateur puisque celui-ci ne présente d'intérêt que si le paramètre est identifiable.

Différents modèles sont étudiés tout au long de la thèse. Nous commençons par un modèle de régression linéaire, puis nous étendons la méthode aux régressions non-linéaires et aux modèles structurels. Cette thèse se compose de trois essais.

Nous étudions dans le premier essai la construction de tests et de régions de confiance dans un modèle de régression linéaire sur la médiane. Nous supposons que le processus d'erreur est de médiane nulle conditionnellement aux variables explicatives et à son propre passé sans imposer de restriction paramétrique supplémentaire sur sa distribution. Celle-ci peut être non gaussienne, hétéroscédastique ou bien présenter une dépendance sérielle de forme arbitraire, ce qui inclut les processus ARCH, GARCH et de volatilité stochastique. Seule est exclue dans un premier temps la dépendance linéaire. La transformation par les signes des résidus contraints permet de définir des statistiques de test dont la distribution ne dépend pas de paramètres de nuisance et est aisément simulable quelle que soit la taille de l'échantillon. La méthode des tests de Monte Carlo ainsi que celle des projections nous permettent tour à tour de construire des tests simultanés exacts et des régions de confiance pour le vecteur de paramètres, puis des tests et des régions de confiance valides pour n'importe quelle transformation possiblement non linéaire, et ce, quelle que soit la taille de l'échantillon.

En revanche, les statistiques de signes que nous utilisons ne sont plus pivotales lorsque le processus d'erreur est linéairement dépendant (cas d'un ARMA stationnaire, par exemple). La matrice de variance asymptotique constitue dans ce cas un paramètre de nuisance. Les méthodes HAC standard nous permettent de le corriger asymptotiquement. La procédure de Monte Carlo développée précédemment est alors asymptotiquement valide sous des hypothèses d'existence de moment et de densité plus faibles que les méthodes asymptotiques habituelles. De plus, elle ne requiert pas d'approximer des paramètres inconnus (ce qui, au contraire, constitue une des principales difficultés des méthodes des noyaux par exemple).

Les études par simulation suggèrent qu'elle est plus performante que les méthodes asymptotiques habituelles pour des processus très hétérogènes ou lorsque la taille de l'échantillon est petite. Cette méthode est donc particulièrement adaptée à l'étude des données financières qui sont souvent très hétéroscédastiques ainsi qu'aux analyses qui s'appuient sur un faible nombre d'observations (séries temporelles, études inter-régionales, données d'enquête, ...).

L'approche inférentielle basée sur les tests permet de mettre en avant d'autres outils moins communément utilisés en économétrie. Dans le deuxième essai, nous reprenons et étudions, la notion de distribution de confiance associée à une statistique de test qui est une réinterprétation des distributions fiducielles de Fisher [Fisher (1930), Efron (1998), Schweder et Hjort (2002)]. La fonction p-value qui en découle associe un *degré de confiance* à chaque valeur testée du paramètre, étant donnée la réalisation des données. Distributions de confiance et fonctions p-value constituent un équivalent fréquentiste aux distributions *a posteriori* bayésiennes. Elles résument les résultats des tests et en donnent une illustration graphique. La distribution de confiance est pourtant rarement utilisée en économétrie car elle ne se définit aisémment que dans le cas d'un paramètre réel et requiert l'utilisation d'une statistique pivotale. La fonction p-value peut, elle, être étendue au cas d'un paramètre multidimensionnel. Notre objectif est d'étendre ces notions au cas multidimensionnel dans le contexte d'une régression sur la médiane. La transformation par les signes nous permet de construire des statistiques pivotales sans recourir à des hypothèses paramétriques. Nous calculons des fonctions p-value simulées à partir de tests de Monte Carlo, puis des versions projetées pour chaque composante individuelle du paramètre. Celles-ci donnent à la fois une illustration graphique de l'inférence et du degré d'identification du paramètre. Cependant, comme elles s'appuient sur des statistiques discrètes, nous n'avons que des versions approchées des notions initiales.

Le deuxième objectif de cet essai est d'associer un estimateur à la procédure d'inférence. Pour ce faire, nous suivons le principe d'inversion de test de Hodges et Lehmann [Hodges et Lehmann (1963)], et proposons d'utiliser comme estimateur la valeur du paramètre associé au plus haut degré de confiance (soit à la plus forte p-value). Nous montrons que l'estimateur de signe qui en découle est sans biais pour la médiane quand les erreurs sont symétriques et qu'il partage les propriétés d'équivariance de l'estimateur « Least Absolute Deviations (LAD) ». Il est aussi convergent et asymptotiquement normal sous des conditions plus faibles que l'estimateur LAD.

En échantillon fini, les simulations suggèrent qu'il est supérieur à l'estimateur LAD, du point de vue du biais et de l'erreur quadratique moyenne, pour des processus très hétéroscédastiques ou possédant des queues de distribution épaisses.

Le troisième essai porte sur les modèles structurels et non-linéaires. L'échec des méthodes asymptotiques usuelles dans les modèles structurels motive fortement une étude à distance finie. Pourtant, la plupart des procédures disponibles dans la littérature s'appuient sur un modèle paramétrique ou ne sont qu'asymptotiquement justifiées. Dans un cadre non paramétrique, seules les procédures de rang ont été adaptées aux échantillons finis. Ce troisième essai présente une procédure valide quelle que soit la taille de l'échantillon et robuste à l'hétéroscédasticité de forme arbitraire. Nous utilisons une version de la propriété de pivotalité des signes adaptée à un modèle avec instruments. Les tests qui en découlent sont exacts et ne dépendent pas du degré d'identification du paramètre. Ils restent valides en présence d'instruments faibles ou de problème d'identification du paramètre.

L'approche que nous suivons peut aussi s'interpréter en termes de régressions artificielles. Les signes des résidus contraints sont régressés sur des instruments auxiliaires dans l'esprit d'Anderson et Rubin [Anderson et Rubin (1949), Dufour (2003)]. Ce type de procédure a cependant le défaut de perdre de la puissance lorsque beaucoup d'instruments sont utilisés. Ceci pose la question des instruments à inclure dans le modèle en cas de suridentification. Nous étudions deux concepts d'optimalité et proposons d'utiliser la méthode de partage de l'échantillon [« split sample », Dufour et Jasiak (2001)] pour calculer des versions approchées de ces instruments optimaux.

Les simulations montrent que notre approche est supérieure aux méthodes usuelles (y compris celles qui sont robustes à la présence d'instruments faibles) lorsque les erreurs sont non gaussiennes et hétéroscédastiques.

Chapitre 1

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

1. Introduction

The Laplace-Boscovich median regression has received a renewed interest since two decades. This method is known to be more robust than least squares and easily allows for heterogeneous data [see Dodge (1997)]. It has recently been adapted to models involving heteroskedasticity and autocorrelation [Zhao (2001), Weiss (1990)], endogeneity [Amemiya (1982), Powell (1983), Hong and Tamer (2003)], nonlinear functional forms [Weiss (1991)] and has been generalized to other quantile regressions [Koenker and Bassett (1978)]. Theoretical advances on the behavior of the associated estimators have completed this process [Powell (1994), Chen, Linton, and Van Keilegom (2003)]. In empirical studies, partly thanks to the generalization to quantile regressions, new fields of potential applications were born.[1] The recent and fast development of computer technology clearly stimulates interest for these robust, but formerly viewed as too cumbersome, methods.

Linear median regression assumes a linear relation between the dependent variable y and the explanatory variables x. Only a null median assumption is imposed on the disturbance process. Such a condition of identification "by the median" 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 $i.i.d.$ observations, for any sample size. Moments are not empirically meaningful without any further distributional assumptions. This form of non-identification can be eliminated, even in finite samples, by choosing another measure of central tendency, such as the median. Hypotheses on the median of non-normal observations can easily be tested by signs tests [see Pratt and Gibbons (1981)]. In nonparametric setups, one may expect models with median identification to be more appropriate than their 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 estima-

[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 (1999) for an application in Value at Risk issues in finance and Koenker and Hallock (2001), Buchinsky (1998), for exhaustive reviews of this literature.

tors, such as the least absolute deviations (LAD) estimator, are more robust to outliers than usual LS methods and may be more efficient whenever the median is a better measure of location than the mean. This holds for heavy-tailed distributions or distributions that have mass at 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 framework [see Powell (1994)]. However, related inference and confidence intervals remain based on asymptotic normality approximations. This reversal to normal approximate inference is certainly disappointing when so much effort has been made to get rid of parametric models.

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

The starting point is a well known result of quasi-impossibility in the non-parametric statistical literature. Lehmann and Stein (1949) 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), Lehmann (1959)]. This result has two main consequences. First, sign-based methods, which do control the conditional level, are a 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 interest in econometrics, compared to ranks or signed ranks methods. Dufour (1981), Campbell and Dufour (1991, 1995), Wright (2000), derived exact nonparametric tests for different time series models. In a regression context, Boldin, Simonova, and Tyurin (1997) developed inference and estimation for linear models. They presented both exact and asymptotic-based inferences for $i.i.d.$ observations, whereas for autoregressive processes with $i.i.d.$ disturbances, only asymptotic justification was available. Our work is positioned in the following of Boldin, Simonova, and Tyurin (1997). We keep sign-based statistics related to locally optimal sign tests, which are simple quadratic forms and can easily be adapted for estimation. However, we extend their distribution-free properties to allow for a wide array of nonlinear dependent schemes. We propose to conjugate them with projection techniques and Monte Carlo tests to systematically derive exact confidence sets.

The pivotality of the sign-based statistics validates the use of Monte Carlo tests, a technique proposed by Dwass (1957) and Barnard (1963). The Monte Carlo method, adapted to discrete statistics by a tie-breaking procedure [Dufour (2006)], yields exact simultaneous confidence region for β. Then, conservative confidence intervals (CIs) for each component of the parameter (or any real function of the parameter) are obtained by projection [Dufour and Kiviet (1998), Dufour and Taamouti (2005), Dufour and Jasiak (2001)]. Exact CIs as they are valid can be unbounded for nonidentifiable component. That results from the exactness of the method and insures the true value of the component belongs to exact CIs with probability higher than $1 - \alpha$. In practice, computation of bounds of confidence intervals (or confidence sets) requires global optimization algorithms such as simulated annealing [see Goffe, Ferrier, and Rogers (1994)].

Sign-based inference methods constitute an alternative to inference derived from the asymptotic behavior of the well known LAD estimator. 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 (1997b) 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 (1997b) 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) and Fitzenberger (1997a), Buchinsky and J. (1998), "simultaneous equations", Amemiya (1982), Hong and Tamer (2003)]. More recently, authors have been interested in allowing 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. 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 levels can be far from their nominal size. One can find examples of such distortions for time series context in Dufour (1981), Campbell and Dufour (1995, 1997) and for L_1-estimation in Buchinsky (1995), De Angelis, Hall, and Young (1993), Dielman and Pfaffenberger (1988a, 1988b). Inference based on signs constitutes an alternative that does not suffer from these shortcomings.

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 region on β can be derived using Monte Carlo tests. For more general processes that may involve stationary ARMA disturbances, these statistics are no longer pivotal. The serial dependence parameters constitute nuisance pa-

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

rameters. 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 method. For these kinds of processes, we loose the exactness but keep an asymptotic validity. In particular, this asymptotic validity requires less assumptions on moments or the shape of the distribution (such as the existence of a density) than usual asymptotic-based inference. Besides, we do not need to evaluate the disturbance density at zero, which constitutes one of the major difficulties of kernel-based methods. In practice, we derive sign-based statistics from locally most powerful test statistics. We obtain exact simultaneous confidence region and then, conservative confidence intervals for each component or any real function of β by projection techniques. Once again, we stress the fact that sign-based statistics can provide finite-sample inference which is not the case for usual inference theories associated with LAD and other quantile estimators, which rely on their asymptotic distributions.

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. Appendix A contains the proofs.

2. Framework

2.1. Model

We consider a stochastic process $W = \{W_t = (y_t,\ x_t') : \Omega \to \mathbb{R}^{p+1},\ t = 1, 2, \ldots\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$. Let $\{W_t, \mathcal{F}_t\}_{\,t=1,2,\ldots}$ be an adapted stochastic sequence, *i.e.* $\mathcal{F}_t$ is a σ-field in Ω such that $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s < t$ and $\sigma(W_1, \ldots, W_t) \subset \mathcal{F}_t$, where

$\sigma(W_1, \ldots, W_t)$ is the σ-algebra spanned by $W_1, \ldots, W_t$. $W_t = (y_t, x_t')$, where y_t is the dependent variable and $x_t = (x_{t1}, \ldots, x_{tp})'$, a p-vector of explanatory variables . The x_t's may be random or fixed.

We assume that y_t and x_t satisfy a linear model and we shall impose in the following some conditions on the median of the disturbance process:

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

In the following, $y = (y_1, \ldots, y_n)' \in \mathbb{R}^n$ stands for the dependent vector, $X = [x_1, \ldots, x_n]'$ for the $n \times p$ explanatory matrix. $\beta \in \mathbb{R}^p$ is the vector of parameters, and $u = (u_1, \ldots, u_n)' \in \mathbb{R}^n$ the disturbance vector. Moreover, the distribution function of u_t conditional on X is denoted $F_t(.|x_1, \ldots, x_n)$.

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

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

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

We depart from this usual assumption. Indeed, our aim is to develop a framework that is robust to heteroskedasticity of unknown form. From Bahadur and Savage (1956), 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 its distribution is impossible. Such a test has no power. This problem of non-testability can be viewed as a form of non-identification in a wide sense. Unless relatively strong distributional assumptions are made, moments are not empirically meaningful. Thus, if one wants to relax the distributional assumptions, one must choose another measure of central tendency such as the median. The median is in particular well adapted if the distribution of the disturbance process does not possess moments. As a consequence, in this median regression framework, the martingale difference assumption will be replaced by an analogue in terms of median. We define the median-martingale difference or shortly said, *mediangale* that can be stated unconditional or conditional on the design matrix X.

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

$$\mathsf{P}[u_1 < 0] = \mathsf{P}[u_1 > 0] = 0.5,$$

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

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

$$\mathsf{P}[u_1 < 0|X] = \mathsf{P}[u_1 > 0|X] = 0.5,$$

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

Note that the above distributions allow u_t to have a discrete distribution except at zero. If the latter constraint is relaxed, we get that following definition.

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

$$\mathsf{P}[u_1 > 0|X] = \mathsf{P}[u_1 < 0|X],$$

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

The sign operator $s : \mathbb{R} \to \{-1, 0, 1\}$ is defined as

$$s(a) = \mathbf{1}_{[0,+\infty)}(a) - \mathbf{1}_{(-\infty,0]}(a), \ \mathbf{1}_A(a) = \begin{cases} 1, \text{ if } a \in A, \\ 0, \text{ if } a \notin A. \end{cases} \tag{2.2}$$

For convenience, the notation will be extended to vectors. Let $u \in \mathbb{R}^n$ and $s(u)$, 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 exactly 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 as defined before differs from a martingale difference on the signs because of the conditioning upon 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 $\{\mathcal{F}_t = \sigma(W_1, \ldots, W_t, X), \ t = 1, 2, \ldots\}$. Conditional mediangale requires conditioning on the whole process X. We shall see later that asymptotic inference may be available under weaker assumptions, as a classical martingale difference on signs or more generally some mixing concepts 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 $\{u_t, \ t = 1, 2, \ldots\}$ by a quasi-similar hypothesis on a robust transform of this process $\{s(u_t), \ t = 1, 2, \ldots\}$. Below we will see it is relatively easy to deal with a weak mediangale by a simple transformation of the sign operator. To simplify the presentation, we shall focus on the strict mediangale concept. Therefore, our model will rely on the following assumption.

Assumption A1 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 A1 entails:

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

Hence, we are in a median regression context. Our last remark concerns exogeneity. As long as the x_t's are strongly exogenous explanatory variables,[3] the conditional mediangale concept is equivalent to usual martingale difference for signs with respect to $\mathcal{F}_t = \sigma(W_1, \ldots, W_t),\ t = 1, 2, \ldots$.

Proposition 2.5 MEDIANGALE EXOGENEITY. *Suppose* $\{x_t : t = 1, \ldots, n\}$ *is a strongly exogenous process for* β *and*

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

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

Model (2.1) with the Assumption A1 allows for very general forms of the disturbance distribution, including asymmetric, heteroskedastic or dependent ones, as long as conditional medians are 0. We stress that neither density nor moment existence are required, which is an important difference with asymptotic theory. Indeed, what the mediangale concept requires is a form of independence in the signs of the residuals. This extends results in Dufour (1981) and Campbell and Dufour (1991, 1995, 1997).

Asymptotic normality of the LAD estimator is presented in its most general way in Fitzenberger (1997b). It holds under some mixing concepts on $\{s(u_t), \sigma(W_1, \ldots, W_t),\ t = 1, 2, \ldots\}$ and an orthogonality condition between $\{s(u_t),\ t = 1, 2, \ldots\}$ and $\{x_t,\ t = 1, 2, \ldots\}$. However, this requires additional assumptions on moments.[4] With such a choice, testing is necessarily based on approximations (asymptotic or bootstrap). Here, we focus on valid finite-sample inference without any further assumption on the form of the distributions. In order to conduct a fully exact method, we have to consider Assumption A1.

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

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

2.2. Special cases

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

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

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

$$\begin{aligned} u_1 &= \sigma_1(x_1, \ldots, x_{t-1})\varepsilon_1\,, \\ u_t &= \sigma_t(x_1, \ldots, x_{t-1}\,, u_1,\, \ldots\,,\, u_{t-1})\varepsilon_t\,, \quad t = 2,\, \ldots\,,\, n \end{aligned}$$

where $\varepsilon_1,\, \ldots\,,\, \varepsilon_n$ are independent with median 0, $\sigma_1(x_1, \ldots, x_{t-1})$ and $\sigma_t(x_1, \ldots, x_n\,, u_1,\, \ldots\,,\, u_{t-1})$, $t = 2, \ldots, n$ are non-zero with probability one. In time series context, this includes:

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

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

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

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

3. stochastic volatility models with non-Gaussian noises ε_t:

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

The first example is especially relevant for cross-sectional data when procedures are expected to be robust to heteroskedasticity. Other examples present robustness properties to endogenous disturbance variance (or volatility) 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. Note finally that the previous property is more general and does not specify explicitly the functional form of the variance in contrast with an ARCH specification.

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.[5] 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 very misleading. Test levels can be quite far from their nominal size: 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 (1981), Campbell and Dufour (1995, 1997) and Mankiw and Shapiro (1986)]; on inference based on L_1 estimators, see also Buchinsky (1995), De Angelis, Hall, and Young (1993), Dielman and Pfaffenberger (1988a, 1988b). This remark usually

[5]For the estimation theory, the reader is referred to Coudin and Dufour (2005b).

motivates the use of bootstrap procedures. In a sense, bootstrapping (once bias corrected) is a way to make approximation closer by introducing artificial observations. However, the bootstrap still relies on approximations and in general there is no guarantee that the level condition is satisfied in finite samples.

Another way to appreciate the nonvalidity of asymptotic methods in finite samples is to recall a 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 (recalled and proved in Pratt and Gibbons (1981), see also Lehmann (1959).

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

$$\mathsf{P}[\textit{Rejecting } H_0 \mid |X_1| , \ldots , |X_n|] \leq \alpha \textit{ under } H_0 . \tag{3.2}$$

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 condition (3.2) has size one. Note that procedures typically designated as "robust to heteroskedasticity" or "HAC" [see White (1980), Newey and West (1987), Andrews (1991), etc.] are not proved to satisfy condition (3.2), so they can have size one for any sample size.

Sign-based procedures do satisfy this condition. Besides, as we will show in the next section, 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 mutually independent equalling 1 with probability $1/2$ and -1 with probability $1/2$. We state more precisely this result in the following proposition. We see also that 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* $(u_1, \ldots, u_n)$ *satisfy a strict mediangale conditional on* $X = [x_1, \ldots, x_n]'$. *Then the variables* $s(u_1), \ldots, s(u_n)$ *are i.i.d. conditional on* X *according to the distribution*

$$\mathsf{P}[s(u_t) = 1 \,|\, x_1, \ldots, x_n] = \mathsf{P}[s(u_t) = -1 \,|\, x_1, \ldots, x_n] = \frac{1}{2}, \quad t = 1, \ldots, n. \tag{3.3}$$

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

From the above proposition, it follows that the residual sign vector of the model constrained to β_0

$$s(y - X\beta) = [s(y_1 - x_1'\beta), \ldots, s(y_n - x_n'\beta)]' \tag{3.4}$$

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

$$T = T\big(s(y - X\beta), X\big) \tag{3.5}$$

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 and exactly controlled. Consider testing

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

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

$$T\big(s(y - \beta_0 X), X\big) \sim T(S_n, X) \tag{3.6}$$

where $S_n = (s_1, \ldots, s_n)$ and $s_1, \ldots, s_n$ are $i.i.d.$ random variables according to a uniform Bernoulli distribution on $\{-1,\ 1\}$. A test with level α rejects the null hypothesis when

$$T\big(s(y - \beta_0 X), X\big) > c_T(X, \alpha) \tag{3.7}$$

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

This method can be extended to error distributions with a mass at zero, $i.e.$,

$$\begin{aligned} &\mathsf{P}[u_1 > 0 \,|\, X] = \mathsf{P}[u_1 < 0 \,|\, X], \\ &\mathsf{P}[u_t > 0 \,|\, X,\ u_1,\ \ldots,\ u_{t-1}] = \mathsf{P}[u_t < 0 \,|\, X,\ u_1,\ \ldots,\ u_{t-1}], \quad t \geq 2\,. \end{aligned} \tag{3.8}$$

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

$$\mathsf{P}[u_t = 0 \,|\, X,\ u_1,\ \ldots,\ u_{t-1}] = p_t(X,\ u_1,\ \ldots,\ u_{t-1}) > 0 \tag{3.9}$$

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

$$\tilde{s}(x,\ V) = s(x) + \big[1 - s(x)^2\big]s(V - 0.5), \quad \text{where} \quad V \sim U(0,\ 1), \tag{3.10}$$

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

$$\mathsf{P}[\tilde{s}(u_t,\, V_t) = +1] = \mathsf{P}[\tilde{s}(u_t,\, V_t) = -1] = \frac{1}{2}\,. \tag{3.11}$$

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 following a* $U(0,\, 1)$ *distribution independent of* u *and* X. *Then the variables* $\tilde{s}_t = \tilde{s}(u_t,\, V_t)$ *are i.i.d. conditional on* X *with the distribution*

$$\mathsf{P}[\tilde{s}_t = 1 \,|\, X] = \mathsf{P}[\tilde{s}_t = -1 \,|\, X] = \frac{1}{2}, \quad t = 1, \ldots, n\,. \tag{3.12}$$

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

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 the exact distribution of those sign-based statistics. 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 provide power against alternatives of interest. Unfortunately, there is no uniformly most powerful test of $\beta = \beta_0$ against $\beta \neq \beta_0$. Hence, different alternatives may be considered. For testing $H_0(\beta_0) : \beta = \beta_0$ against $H_1(\beta_0) : \beta \neq \beta_0$ in model (2.1), we consider test statistics of the following form:

$$D_S(\beta_0, \Omega_n) = s(y - X\beta_0)' X \Omega_n\big(s(y - X\beta_0), X\big) X' s(y - X\beta_0) \tag{4.13}$$

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

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

$$SB(\beta_0) = s(y - X\beta_0)'XX's(y - X\beta_0) = \|X's(y - X\beta_0)\|^2 \tag{4.14}$$

and

$$SF(\beta_0) = s(y - X\beta_0)'P(X)s(y - X\beta_0) = \|X's(y - X\beta_0)\|_M^2 \tag{4.15}$$

where $P(X) = X(X'X)^{-1}X'$. In Boldin, Simonova, and Tyurin (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 $f'(0) = 0$).[6] Their proof can easily be extended to disturbances that satisfy the mediangale property and for which the conditional density at zero is the same $f_t(0|X) = f(0|X),\ \forall t = 1, \ldots, n$.

$SF(\beta_0)$ can be interpreted as a sign analogue of the Fisher statistic. More precisely, $SF(\beta_0)$ is a monotonic transformation of the Fisher statistic for testing $\gamma = 0$ in the regression of $s(y - X\beta_0)$ on X:

$$s(y - X\beta_0) = X\gamma + v. \tag{4.16}$$

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 exactly the $SF(\beta_0)$ statistic. The same

[6] The power function of the locally most powerful sign-based test knows 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 on unbiased tests, Boldin, Simonova, and Tyurin (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 (Ch. 2, pp. 76-86, 1984) , or Gray (Ch. 21, pp. 373-380, 1998) , for a presentation of various curvature notions].

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, Simonova, and Tyurin (1997) 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 function in the vicinity of 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 A1 holds, and the disturbances are heteroskedastic with conditional densities* $f_i(.|X)$, $i = 1, 2, \ldots$, *that are continuously differentiable around zero and such that* $f_i'(0|X) = 0$. *Then, the locally optimal sign test statistic associated with the mean curvature is*

$$\tilde{SB}(\beta_0) = s(y - X\beta_0)'\tilde{X}\tilde{X}'s(y - X\beta_0) \tag{4.17}$$

where

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

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.

These test statistics 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 A1. 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 this asymptotic covariance matrix by $J_n\big(s(y - X\beta_0), X\big)$. We consider

$$\Omega_n\big(s(y - X\beta_0), X\big) = \frac{1}{n}\hat{J}_n\big(s(y - X\beta_0), X\big)^{-1} \tag{4.18}$$

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

$$D_S(\beta_0, \frac{1}{n}\hat{J}_n^{-1}) = \frac{1}{n}s(y - X\beta_0)'X\hat{J}_n^{-1}X's(y - X\beta_0). \tag{4.19}$$

$J_n\big(s(y - X\beta_0), X\big)$ 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. A more elegant solution consists in using the technique of **Monte Carlo tests**, which can be viewed as a finite-sample version of the **bootstrap**.

4.2. Monte Carlo tests

Monte Carlo tests have been introduced by Dwass (1957) and Barnard (1963) and can be adapted to 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).

All the tests presented above are on the same model: given a statistic T, the test rejects the null hypothesis when T is large, *i.e.* when $T \geq c$, where c depends on the level of the test. Moreover, the conditional distribution of T given X is free of nuisance parameters. All ingredients are present to apply Monte Carlo test procedures.

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

$$\hat{p}_N(x) = \frac{N\hat{G}_N(x) + 1}{N + 1} \tag{4.20}$$

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 $\alpha(N+1)$ is an integer, then $\mathsf{P}[\hat{p}_N(T^{(0)}) \leq \alpha] = \alpha$. The level of the test 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. Different procedures have been presented in the literature to decide what to do when ties occur. 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).

Here, we use a randomized tie-breaking procedure for evaluating empirical survival functions in case of discrete statistics. The latter is based on replacing the usual order relation by a lexicographic order relation can be used [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:

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

This leads to the following p-value function:

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

where

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

Then

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

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

Here, we consider testing $H_0(\beta_0)$ in (2.1) under a mediangale assumption on the errors using a statistic of the form $DS(\beta, \Omega_n)$. Take, for example, $SF(\beta)$. After computing $SF^{(0)} = SF(\beta_0)$ from the data, we choose N the number of replicates, such that $\alpha(N+1)$ is an integer, where α is the desired level. Then, we generate N replicates $SF^{(j)} = S^{(j)\prime}X(X'X)^{-1}X'S^{(j)}$ where $S^{(j)}$ is a realization of a n-vector of independent Bernoulli random variables, and we compute $\tilde{p}_N[SF^{(0)}]$. 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 components. More generally, we can obtain conservative confidence sets for any transformation $g(\beta)$ where g can be any kind of real function, including nonlinear ones.

Obviously, obtaining a continuous grid of $\mathbb{R}^p$ is not realistic. We will instead require **global optimization search algorithms**.

5.1. Confidence sets and conservative confidence intervals

Projection techniques yield finite-sample valid confidence intervals and confidence sets for general functions of the parameter β.[7] 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

$$\beta \in C_{1-\alpha}(\beta) \Longrightarrow g(\beta) \in g\big(C_{1-\alpha}(\beta)\big), \tag{5.1}$$

we have:

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

[7]For examples of use in different settings and for further discussion, the reader is referred to Dufour (1990, 1997), Abdelkhalek and Dufour (1998), Dufour and Kiviet (1998), Dufour and Jasiak (2001), Dufour and Taamouti (2005).

Thus, $g(C_{1-\alpha}(\beta))$ is a conservative confidence set for $g(\beta)$. If $g(\beta)$ is scalar, the interval (in the extended real numbers)

$$I_g[C_{1-\alpha}(\beta)] = \left[\inf_{\beta \in C_{1-\alpha}(\beta)} g(\beta)\,,\ \sup_{\beta \in C_{1-\alpha}(\beta)} g(\beta)\right]$$

has level $1-\alpha$:

$$\mathsf{P}\left[\inf_{\beta \in C_{1-\alpha}(\beta)} g(\beta) \leq g(\beta) \leq \sup_{\beta \in C_{1-\alpha}(\beta)} g(\beta)\right] \geqslant 1-\alpha. \tag{5.2}$$

Hence, to obtain valid conservative confidence intervals for the component β_k of the β parameter in the model (2.1) under mediangale Assumption A1, 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:

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

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

where $\tilde{p}_N$ is computed as proposed in the previous section, using N replicates $SF^{(j)}$ of the statistic SF under the null hypothesis. This can be done easily in practice with a global search optimization algorithm, like **simulated annealing** [see Goffe, Ferrier, and Rogers (1994), and Press, Teukolsky, Vetterling, and Flannery (2002)]. The method allows one to perform tests for general hypotheses and to derive confidence sets. In the case of multiple tests, an arbitrary number of hypotheses can be tested without ever loosing control of the overall level: rejecting at least one true null hypothesis will not exceed the specified level α.

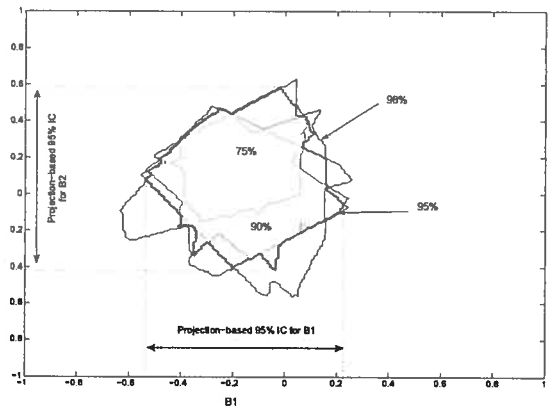


Figure 1. Confidence regions provided by SF-based inference.

5.2. Numerical illustration

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

$$y_t = \beta_0 + \beta_1 x_t + u_t\,, \quad t = 1, \ldots, n, \tag{5.3}$$

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

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 bigger 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 obtained confidence regions 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 outliers.

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]

6. Asymptotic theory

This section is dedicated to asymptotic results. We point out that the mediangale Assumption A1 can be seen as too restrictive and excludes some common processes whereas usual asymptotic inference still can be conducted on them. We relax Assumption A1 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 a HAC-corrected version of the sign-based statistic $D_S(\beta_0, \frac{1}{n}\hat{J}_n^{-1})$ in (4.19) 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 A2 MIXING. $\{(x_t', u_t)\}_{t=1,2,\ldots,}$ *is α-mixing of size $-r/(r-2)$ with $r > 2$.*[8]

Assumption A3 MOMENT CONDITION. $\mathsf{E}[s(u_t)x_t] = 0,\ \forall t = 1, \ldots, n,\ \forall n \in \mathbb{N}$.

Assumption A4 BOUNDEDNESS. $x_t = (x_{1t}, \ldots, x_{pt})'$ *and* $E|x_{ht}|^r < \Delta < \infty,\ h = 1, \ldots, p,\ t = 1, \ldots, n,\ \forall n \in \mathbb{N}$.

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

Assumption A5 NON-SINGULARITY. $J_n = \mathrm{var}[\frac{1}{\sqrt{n}}\sum_{t=1}^{n} s(u_t)x_t]$ *is uniformly positive definite.*

Assumption A6 CONSISTENT ESTIMATOR OF J_n. $\Omega_n(\beta_0)$ *is symmetric positive definite uniformly over* n *and* $\Omega_n - \frac{1}{n}J_n^{-1} \to_p 0$.

Theorem 6.1 ASYMPTOTIC DISTRIBUTION OF STATISTIC SHAC. *In model* (2.1), *with Assumptions A2- A6, we have, under* H_0,

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

Corollary 6.2 *In model* (2.1), *suppose the mediangale Assumption A1 and boundedness Assumption A4 are fulfilled. If* $X'X/n$ *is positive definite uniformly over* n *and converges in probability to a definite positive matrix, then, under* H_0,

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

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

6.2. Asymptotic validity of Monte Carlo tests

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

6.2.1. Generalities

Let us consider a parametric or semiparametric model $\{M_\beta,\ \beta \in \Theta\}$, where the parameter β is identified. 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 $F_n(x)$ and $G_n(x)$ is the corresponding survival function. We suppose that $F_n(x)$ converges almost everywhere to a distribution function $F(x)$. Let $G(x)$ be the corresponding survival function. In Theorem 6.3, we show the following: if a series of conditional survival functions $\tilde{G}(x|X_n(\omega))$ given

$X(\omega)$ satisfies

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

where G does not depend on the realization $X(\omega)$, then $G_n(x)$ can be approximated by $\tilde{G_n}(x|X_n(\omega))$. Consequently, $\tilde{G_n}(x|X_n(\omega))$ can be seen as an approximation of $G_n(x)$ or a *pseudo* survival function of $c_n S_n(\beta_0)$. Note that $G(x)$ can depend on some parameters of the distribution of X.

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 series of survival functions such that

$$\tilde{G_n}(x|X_n(\omega)) \underset{n\to\infty}{\to} G(x) \quad \textit{with probability one.}$$

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 the null hypothesis is

$$\tilde{G}_n^{(N)}(x, n, S_n^{(0)}, S_n(N), U(N+1)) = 1 - \frac{1}{N}\sum_{j=1}^{N} u(x - c_n S_n^{(j)}) + \frac{1}{N}\sum_{j=1}^{N} \delta(c_n S_n^{(j)} - x) u(U^{(j)} - U^{(0)}) \tag{6.2}$$

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

$$\tilde{p}_n^{(N)}(x) = \frac{N\tilde{G}_n^{(N)}(x) + 1}{N+1}. \tag{6.3}$$

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 $H_0(\beta_0)$,*

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

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

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

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.

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

These results can be applied to 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 A2- A6 hold and consider the testing problem

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

Let $D_S(\beta, \hat{J}_n^{-1})$ be the test statistic as defined in (4.19).

- Observe $SF^{(0)} = D_S(\beta_0, \hat{J}_n^{-1})$. Draw N replicates of sign vector as if the n observations were independent. The n components of the sign vectors are independent and drawn from a $B(1, .5)$ distribution.
- Compute $(SF^{(1)},\ SF^{(2)}, \ldots, SF^{(N)})$, the N *pseudo* replicates of $D_S(\beta_0, X'X^{-1})$ under the null hypothesis. 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 (6.3).
- 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 $1 - \alpha$. We reject $\mathcal{H}_0$ if $\tilde{p}_n^{(N)}(\beta_0) \leq \alpha$.

Remark that, 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 (1997b) and Weiss (1991). Besides, asymptotic variance of the LAD estimator involves the conditional density at zero of the disturbance process $\{u_t;\ t = 1, 2, \ldots\}$ as unknown nuisance parameter. The approximation and estimation of asymptotic covariance matrix 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 $D_S(\beta, (X'X)^{-1})$ and $D_S(\beta, \hat{J}_n^{-1})$ when a correction is needed for linear serial dependence. We consider a set of general 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:

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

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, $\{u_t,\ t = 1, 2 \ldots\}$ is $i.i.d.$ or depends on the explanatory variables and its past values in a *multiplicative* heteroskedastic or dependent and stationary way,

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

In those cases, the error term constitutes a strict conditional mediangale given X (see Assumption A1). 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 8-10, 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 11 and 12 illustrate two kinds of second-order nonstationary disturbances. As we noted previously, sign-based inference does not require stationary assumptions in contrast with asymptotic tests derived from CLT.

More precisely, cases 1 and 2 present $i.i.d.$ normal observations without and with conditional heteroskedasticity. Case 3 involves outliers in the error term. This can be seen as an example of measurement error in the observed y. Cases 4 and 5 involve other het-

Table 2. Simulated models.

CASE 1:	Normal HOM:	$(x_{2,t}, x_{3,t}, u_t)' \overset{i.i.d}{\sim} \mathcal{N}(0, I_3),\ t = 1, \ldots, n$
CASE 2:	Normal HET:	$(x_{2,t}, x_{3,t}, \tilde{u}_t)' \overset{i.i.d}{\sim} \mathcal{N}(0, I_3)$ $u_t = min\{3, max[0.21, \lvert x_{2,t}\rvert]\} \times \tilde{u}_t,\ t = 1, \ldots, n$
CASE 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, $t = 1, \ldots, n$.
CASE 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, ϵ_t, independent, $t = 1, \ldots, n$.
CASE 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, $t = 1, \ldots, n$.
CASE 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, $t = 1, \ldots, n$.
CASE 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, $t = 1, \ldots, n$.
CASE 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, $t = 1, \ldots, n$.
CASE 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.
CASE 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.
CASE 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.
CASE 11:	Nonstat. GARCH(1,1):	$(x_{2,t}, x_{3,t}, \epsilon_t)' \overset{i.i.d.}{\sim} \mathcal{N}(0, I_3),\ t = 1, \ldots, n$, $u_t = \sigma_t \epsilon_t,\ \sigma_t^2 = 0.8u_{t-1}^2 + 0.8\sigma_{t-1}^2$.
CASE 12:	Exp. Var.:	$(x_{2,t}, x_{3,t}, \epsilon_t)' \overset{i.i.d.}{\sim} \mathcal{N}(0, I_3), u_t = exp(.2t)\epsilon_t$.

eroskedastic schemes with stationary GARCH and stochastic volatility disturbances. Case 6 is a very unbalanced design matrix (where the LAD estimator performs poorly). Case 6 BIS combines the previous unbalanced scheme in the design matrix with heteroskedastic disturbances. Case 7 is an example of heavy-tailed errors (Cauchy). Cases 8, 9 and 10 illustrate the behavior of sign-based inference when the error term involves linear dependence at different levels. Finally, cases 11 and 12 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 1-2, 8-10 have been used by Fitzenberger (1997b) to study the performance of block bootstrap (MBB).

7.1. Size

We first study level distortions. We consider the testing problem:

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

We compare exact and asymptotic tests based on $SF = D_S(\beta, (X'X)^{-1})$ and $SHAC = D_S(\beta, \hat{J}_n^{-1})$, where $\hat{J}_n^{-1}$ is estimated by a Bartlett kernel, with various asymptotic tests. Wald and LR-type tests are considered. We consider Wald tests based on the OLS estimate with 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),[9] Bartlett kernel covariance estimator with automatic bandwidth parameter [Powell (1984), Buchinsky (1995)] (BT), design matrix bootstrap centering around the sample estimate (DMB) [Buchinsky (1998)], moving block bootstrap centering around the sample estimate (MBB) [Fitzenberger (1997b)][10]. Finally, we also consider the likelihood ratio statistic (LR) assuming $i.i.d.$ disturbances with an OS estimate of the error

[9]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)].

[10]The block size is 5.

density [Koenker and Bassett (1982)]. Appendix C contains the formulas of the compared estimators and test statistics.

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. The number of replicates for the bootstrap and the Monte Carlo sign-based method is the same, $i.e.$ $N = 2999$. All bootstrapped samples are of size $n = 50$. We simulate $M = 5000$ random samples to evaluate the levels of these tests. For both sign-based statistics, we also report the asymptotic level whenever processes are stationary.

Table 3 contains models where the mediangale condition A1 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 4, 6 BIS). The asymptotic versions of sign-based tests suffer from the same underrejection than other asymptotic tests, suggesting that, for small samples ($n = 50$), the distribution of the test statistic is really far from its asymptotic limit. Hence, the sign-based method that deals directly with this distribution has clearly an advantage on asymptotic methods. When the dependence in the disturbance process is highly nonlinear (Case 6 BIS), the BT method based on a kernel estimation of the LAD asymptotic covariance matrix is not reliable anymore.

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

*: cases when mediangale condition holds.

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

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						
^{a}**CASE 8**: $\rho_\epsilon = .5,\ \rho_x = 0$, HOM	.126 -	.022 *.019***	.171	.124	.118	.085	.151	.201	.240	212
^{a}**CASE 9**: $\rho_\epsilon = \rho_x = .5$, HET	.218 -	.026 *.017***	.440	.131	.097	.108	.308	.407	.328	.276
^{a}**CASE 10**[11]: $\rho_\epsilon = .9,\ \rho_x = 0$, HOM	.521 -	.012 *.003***	.553	.516	.339	.355	.551	.649	.677	.534

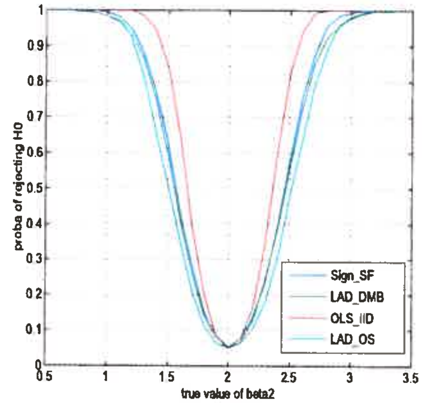
^{a}: cases when mediangale condition fails.
**: sizes using asymptotic critical values based on $\chi^2(3)$.

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 underrejection. 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 overrejections of asymptotic Wald tests based on HAC estimation of the asymptotic covariance matrix when the data set involves a small number of observations. These results suggest, in a sense, that when the data suffer from both a small number of observations and linear dependence, the first problem to solve is the finite-sample distortion, which is not what is usually done.

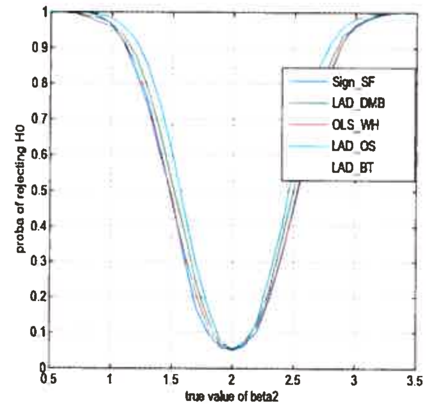
7.2. Power

Then, we illustrate the **power** of these tests. We are particularly interested in comparing the sign-based inference to kernel and bootstrap methods. Others methods may not be reliable even in terms of level. We consider the simultaneous hypothesis H_0 as before. The true process is obtained by fixing β_1 and β_3 at the tested value, i.e $\beta_1 = 1$ and $\beta_3 = 3$, and letting vary β_2. Simulated power is given by a graph with β_2 in abscissa. The power functions presented here (figures 2, 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 consequently loose power.

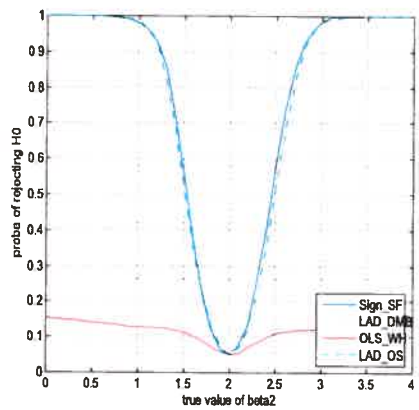
Sign-based inference has a totally comparable power performance with usual methods in cases 1, 2, 3, 8 with the advantage that the level is exactly controlled for any sample size, which leads to great difference in small samples. In very heteroskedastic cases (4, 5, 11, 12), sign-based inference greatly dominates other methods: levels are exactly controlled and power functions largely exceed others, even other methods that are size-corrected with locally adjusted levels. Any HAC correction has only an asymptotic justification. In the presence of linear serial dependence, the Monte Carlo test based on $D_S(\beta, \hat{J}_n^{-1})$ does not exactly control the level in theory for a given sample size. However, it is still asymptotically valid and seems to lead to good power performance, along with a better size control. Only for very high autocorrelation (close to unit root process), the sign-based inference is not adapted anymore.



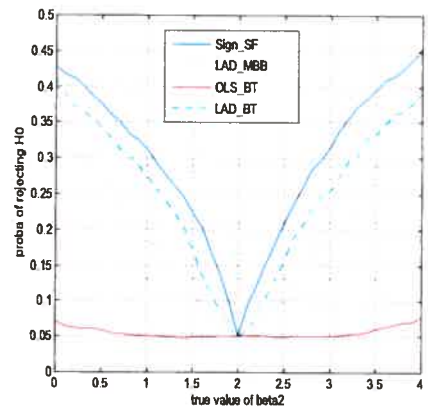
(a) CASE 1: normal



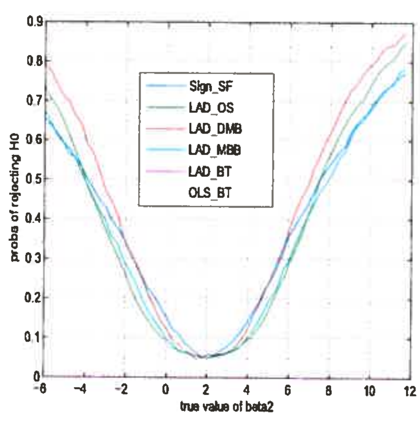
(b) CASE 2: normal HET



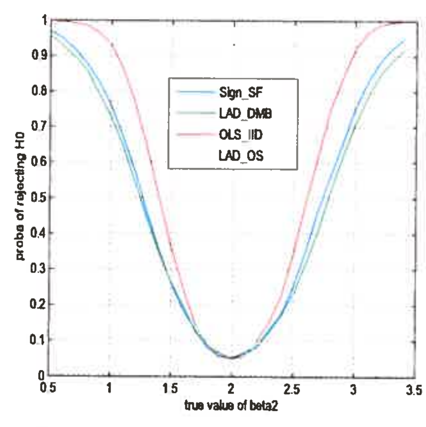
(c) CASE 3: outliers



(d) CASE 4: stationary GARCH



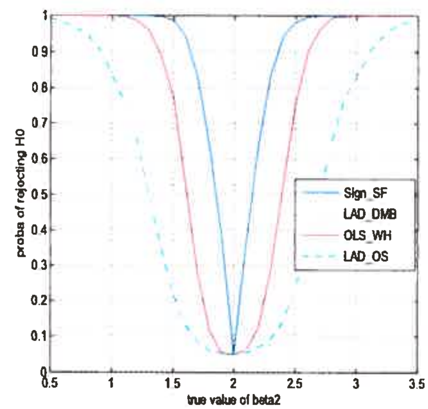
(e) CASE 5: stochastic Volatility



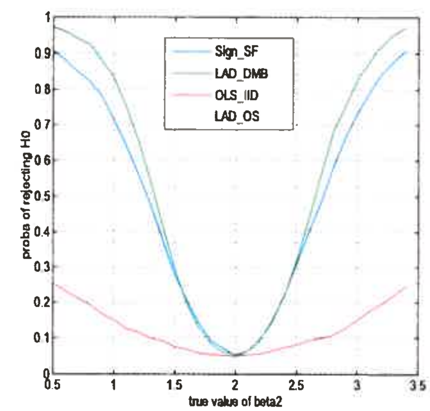
(f) CASE 6: DEB

Sign: $SF = D_S(\beta, X'X^{-1})$, $SHAC = D_S(\beta, \hat{J}_n^{-1})$
LAD/OLS: DMB = design matrix bootstrap, MBB = moving block bootstrap
BT = Bartlett kernel, IID = homoskedastic, WH = White correction, OS = order statistic.

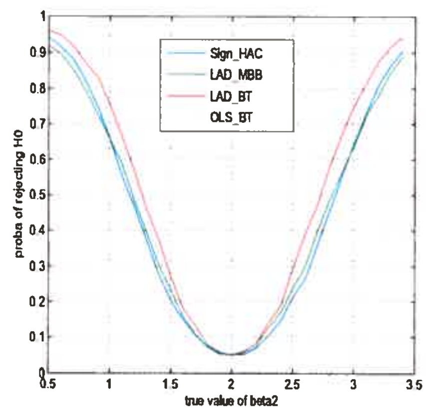
Figure 2. Power functions (level corrected) (1).



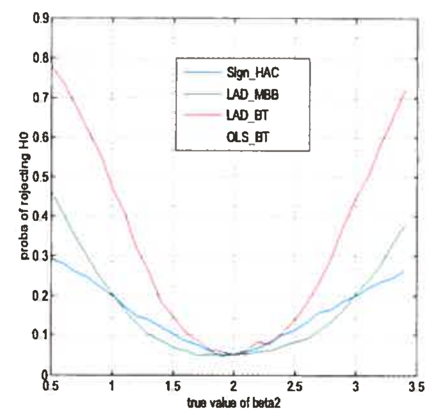
(a) CASE 6BIS: DEB+HET



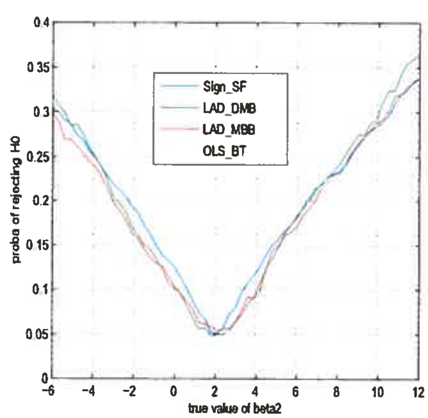
(b) CASE 7: Cauchy



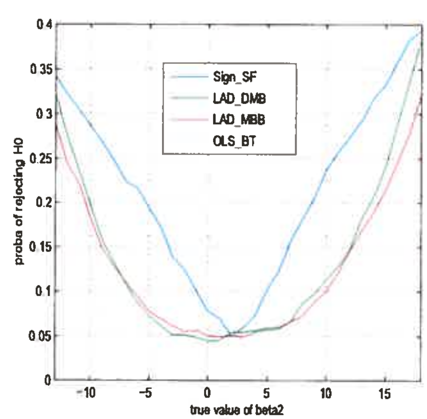
(c) CASE 9: $AR(1)$ HET, $\rho_x = \rho_u = .5$



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



(e) CASE 11 : Nonstationary GARCH



(f) CASE 12: exponential variance

Sign: $SF = D_S(\beta, X'X^{-1})$, $SHAC = D_S(\beta, \hat{J}_n^{-1})$
LAD/OLS: DMB = design matrix bootstrap, MBB = moving block bootstrap
BT = Bartlett kernel, IID = homoskedastic, WH = White correction, OS = order statistic.

Figure 3. Power functions (level corrected) (2).

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 the width of those confidence intervals. 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 the means and the empirical standard deviations of those widths. We only consider the stationary examples. In the nonstationary cases, inference based on t-statistics may not mean anything. In Table 5, we report average width of confidence intervals for each β_k and coverage probabilities. 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 that do not have this feature can even be built.

Table 5. Width of confidence intervals.

$y_t = x_t\beta + u_t,\quad t = 1,\dots,T.$ $T = 50$ $(\beta_1,\beta_2,\beta_3) = (1,2,3)$		*Proj. Based Sign* FS β_1	FS β_2	FS β_3	SHAC β_1	SHAC β_2	SHAC β_3	*t-stat. of LAD* DMB β_1	DMB β_2	DMB β_3	MBB β_1	MBB β_2	MBB β_3	BT β_1	BT β_2	BT β_3
***CASE 1**:	av. spread	1.29	1.52	1.40	1.16	1.36	1.02	.81	.90	.89	.79	.88	.85	.82	.88	.87
$\rho_\epsilon = \rho_x = 0$, HOM	*(st. dev.)*	*(.21)*	*(.27)*	*(.29)*	*(.14)*	*(.28)*	*(.29)*	*(.23)*	*(.21)*	*(.22)*	*(.21)*	*(.24)*	*(.24)*	*(.15)*	(.19)	*(.22)*
	cov. lev.	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**.97**	**.97**	**.97**	**.95**	**.96**	**.95**	**.97**	**.96**	**.96**
***CASE 2**:		.76	1.43	.74	.66	1.26	.48	.43	.94	.43	.42	.90	.41	.50	.92	.50
$\rho_\epsilon = \rho_x = 0$, HET		*(.14)*	*(.29)*	*(.17)*	*(.15)*	*(.28)*	*(.18)*	*(.09)*	*(.24)*	*(.11)*	*(.10)*	*(.27)*	*(.12)*	*(.11)*	*(.29)*	*(.11)*
		1.0	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**.98**	**.97**	**.99**	**.97**	**.95**	**.97**	**.99**	**.95**	**.99**
***CASE 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**
***CASE 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
St. GARCH(1,1):		*(101)*	*(118)*	*(122)*	*(100)*	*(115)*	*(117)*	*(64.6)*	*(74.6)*	*(61)*	*(76.7)*	*(82.6)*	*(84)*	*(70.3)*	*(76.9)*	*(78)*
		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**
***CASE 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**
***CASE 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**
***CASE 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**
^{a}**CASE 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
$\rho_\epsilon = .5,\ \rho_x = 0$, HOM		*(.30)*	*(.32)*	*(.30)*	*(.38)*	*(.31)*	*(.28)*	*(.25)*	*(.26)*	*(.24)*	*(.34)*	*(.26)*	*(.23)*	*(.36)*	*(.23)*	*(.21)*
		.99	1.0	1.0	**.99**	**1.0**	**.99**	**.86**	**.98**	**.99**	**.90**	**.97**	**.97**	**.91**	**.96**	**.95**
^{a}**CASE 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
$\rho_\epsilon = \rho_x = .5$, HET		*(.31)*	*(.40)*	*(.40)*	*(.41)*	*(.51)*	*(.35)*	*(.17)*	*(.33)*	*(.24)*	*(.24)*	*(.42)*	*(.26)*	*(.33)*	*(.55)*	*(.36)*
		1.0	**.99**	**1.0**	**.98**	**.97**	**.94**	**.93**	**.88**	**.98**	**.95**	**.89**	**.98**	**.97**	**.83**	**.97**
^{a}**CASE 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
$\rho_\epsilon = .9,\ \rho_x = 0$, HOM		*(.84)*	*(.82)*	*(.95)*	*(1.06)*	*(.68)*	*(.96)*	*(.57)*	*(.56)*	*(.61)*	*(1.00)*	*(.60)*	*(.63)*	*(1.46)*	*(.47)*	*(.61)*
		.68	**.99**	**1.0**	**.74**	**1.0**	**.99**	**.47**	**.95**	**.98**	**.66**	**.97**	**.98**	**.71**	**.87**	**.91**

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 Valéry and Dufour (2004) 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, then converted in price movements, $y_t = 100[\log(SP_t) - \log(SP_{t-1})]$ and adjusted for systematic calendar effects. We consider a model involving a constant and a drift:

$$y_t = a + bt + u_t, \quad t = 1, \ldots, 16127; \tag{8.3}$$

and 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 Breush-Pagan tests for heteroskedasticity both reject homoskedasticity at 1%.[12].

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 dif-

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

ferent 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.[13]

In Table 6, we report 95% confidence intervals for a and b obtained by various methods: finite-sample sign-based method (for sign-based statistics FS and $SHAC$ involving a HAC correction); LAD and OLS with different estimates of their asymptotic covariance matrices (order statistic, bootstrap, kernel...). If the mediangale Assumption A1 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 greatly 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 seem really 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.

[13] 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.

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]

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, finite-sample sign-based tests remains asymptotically valid such as Wald methods.

However, they are also corrected for the sample size and yield to 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

$$\begin{aligned}(1/T)\ln(y_{i,t}/y_{i,t-T}) &= a - [\ln(y_{i,t-T})] \times [(1-e^{-\beta T})/T] + x_i'\delta + \epsilon_i^{t,T}, \qquad (8.4)\\ & i = 1,\dots,48,\ T = 8, 10 \text{ or } 20,\\ & t = 1900, 1920, 1930, 1940, 1950, 1960, 1970, 1980, 1988.\end{aligned}$$

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 when the data suffers from heteroskedasticity, departure from normality, presence of outliers or observations with possibly high influence.

Therefore, we first study whether such problems are present. Regression diagnostics are summarized in Table 8 in the Appendix B and presented in details in Figures 4-21. One can notice that departures from a normal standard case are present in most periods.[14] For example, clues pointing to high influential observations, heteroskedasticity and non-normality of the residuals exist for the basic equation in the 1880-1900 period. Only, the

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

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:

$$(1/T)\ln(y_{i,t}/y_{i,t-T}) = a + \gamma[\ln(y_{i,t-T})] + x_i'\delta + \epsilon_i^{t,T} \tag{8.5}$$

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 95%-confidence intervals (Table 7).

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 8. 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.

Table 7. Regressions for personal income across U.S. States, 1880-1988.

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.

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. First, we test the presence of a drift on the Standard and Poor's Composite Price Index (S&P), for the whole period 1928-1987 and for various shorter subsamples. Secondly, 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

A.1. Proof of Proposition 2.5

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 $s_t = s(u_t)$.

A.2. Proof of Proposition 3.2

Consider the vector $[s(u_1), s(u_2), \ldots, s(u_n)]' \equiv (s_1, s_2, \ldots, s_n)'$. From Assumption A1, 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:

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

Hence, conditional on X, $s_1, s_2, \ldots, s_n$ are distributed like n i.i.d random variables with distribution:

$$\mathsf{P}[s_t = 1] = \mathsf{P}[s_t = -1] = \frac{1}{2}, \ t = 1, \ldots, n.$$

A.3. Proof of Proposition 3.3

Consider model (2.1) with $\{u_t,\ t = 1, 2, \ldots\}$ being a weak conditional mediangale given X. Let show that $[\tilde{s}(u_1), \tilde{s}(u_2), \ldots, \tilde{s}(u_n)]$ can have the same role in Proposition **3.2** as $[s(u_1), s(u_2), \ldots, s(u_n)]$ under Assumption A1. From equation (3.10), we have:

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

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

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

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

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

Substituting (A.2) into (A.1) yields

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

In a similar way,

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

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

A.4. 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 A1, the locally optimal sign-based test (conditional on X) of $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$ is well defined. Among tests with a given confidence 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 first to compute a Taylor expansion at order one around zero and then to identify the terms. Under the mediangale Assumption A1, we have

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

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

Assuming the existence of continuous densities at zero, a Taylor expansion at order one entails:

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

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

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

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

Therefore, the required test has the form

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

or equivalently,

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

where c_α and c'_α are defined by the significance level.

When the disturbances have a common conditional density at zero, $f(0|X)$, we find the results of Boldin, Simonova, and Tyurin (1997). The locally optimal sign-based test is given by

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

The statistic does not depend on the conditional density evaluated at zero.

When $p > 1$, we need an extension of the notion of slope around zero for a multidimensional parameter. Boldin, Simonova, and Tyurin (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,

$$\left.\frac{dP_\beta(W_\alpha)}{d\beta}\right|_{\beta=0}, \tag{A.13}$$

As $f_i'(0) = 0, \ \forall i$, the behavior of the power function around zero is characterized by the quadratic term of its Taylor expansion

$$\beta'\frac{1}{2}\left(\frac{d^2P_\beta(W_\alpha)}{d\beta^2}\right)\beta = \frac{1}{2^{n-2}}\sum\sum_{1\leq i\neq\, j\leq n}[f_i(0|X)s_i\beta' x_i][f_j(0|X)s_j x_j'\beta]. \tag{A.14}$$

The locally most powerful sign-based test in the sense of the mean curvature maximizes the mean curvature which is, by definition, proportional to the trace of $\left.\frac{d^2P_\beta(W_\alpha)}{d\beta^2}\right|_{\beta=0}$; 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.14), we find (after some computations) that

$$\text{tr}\left(\left.\frac{d^2P_\beta(W_\alpha)}{d\beta^2}\right|_{\beta=0}\right) = \sum\sum_{1\leq i\neq\, j\leq n} f_i(0|X)f_j(0|X)s_is_j\sum_{k=1}^{p} x_{ik}x_{jk}. \tag{A.15}$$

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

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

Hence, the locally optimal sign-biased test in the sense developed by Boldin, Simonova, and Tyurin (1997) for heteroskedastic signs, is

$$W_\alpha = \{s : s'(y)\tilde{X}\tilde{X}'s(y) > c_\alpha'\}. \tag{A.17}$$

Another quadratic test statistic convenient for large-sample evaluation is obtained by standardizing by $\tilde{X}'\tilde{X}$:

$$W_\alpha = \{s : s'(y)\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'s(y) > c_\alpha'\}. \tag{A.18}$$

A.5. 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 A5, $V_n^{-1/2}$ 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 A2 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 A3 and A4 imply

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

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

Note also that

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

The mixing property of $\mathcal{Z}_{nt}$ and equations (A.19)-(A.21) allow one to apply a central limit theorem [see White (2001), Theorem 5.20] that yields

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

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

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

Finally, Assumption A6 states that Ω_n is a consistent estimate of V_n^{-1}. Hence,

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

and,

$$n^{-1} s'(y - X\beta_0) X \Omega_n X' s(y - X\beta_0) \to \chi^2(p). \tag{A.25}$$

A.6. Proof of Corollary 6.2

Let $\mathcal{F}_t = \sigma(y_0, \ldots, y_t, x'_0, \ldots, x'_t)$. When the mediangale Assumption A1 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,

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

and $X'X/n$ is a consistent estimate of $E(X'X/n)$. Theorem **6.1** yields $SF(\beta_0) \to \chi_2(p)$.

A.7. Proof of Theorem 6.3

First, we prove the following lemma **A.1** which will be needed in the proof of Theorem **6.3**.

Lemma A.1 *Let $(F_n)_{n\in\mathbb{N}}$ and F be right continuous distribution functions. Suppose that,*

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

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

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

Proof: Suppose reversely that there exist $\epsilon > 0$, a sequence $\{n_k,\ k \in \mathbb{N}\}$ of integers tending to $+\infty$, and a real sequence $\{x_k,\ k \in \mathbb{N}\}$, such that for all k, $|F_{n_k}(x_k) - F(x_k)| \geq \epsilon > 0$. If $\{x_k\}$ is not a convergent sequence, consider instead a convergent subsequence. This can be done as $\mathbb{R} \cup \{-\infty, +\infty\}$ is compact. Cases when $x_k \to \infty$ can be excluded as $F_n(+\infty) = 1 = F(+\infty)$ and $F_n(-\infty) = 0 = F(-\infty)$ by the definition of distribution functions. Hence, without loss of generality, we can choose $\{x_k\} \to \xi$ where $-\infty < \xi < +\infty$.

Let us consider two sequences $\{r^a_m\}$ and $\{r^b_m\}$ tending to ξ and such that $r^a_m < \xi < r^b_m$. For sufficiently large k, we face the following cases,
if $\{x_k\}$ is increasing and $x_k < \xi$:

$$\begin{aligned}\epsilon\ &\leq F_{n_k}(x_k) - F(x_k) \leq F_{n_k}(\xi^-) - F(r^a_m)\\ &\leq F_{n_k}(\xi^-) - F_{n_k}(\xi) + F_{n_k}(r^b_m) - F(r^b_m) - F(r^a_m);\end{aligned}$$

if $\{x_k\}$ is increasing and $x_k < \xi$:

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

if $\{x_k\}$ is decreasing and $x_k \ge \xi$:

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

and if $\{x_k\}$ is decreasing and $x_k \le \xi$:

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

In each case, for fixed k, m can be chosen such that r_m^b and r_m^a are arbitrarily close to ξ. Then, using right continuity properties of F and F_n, the right hand member of each chain of inequalities does not exceed a quantity that tends to zero as $n_k \to \infty$. Thus a contradiction is obtained. We conclude on the uniform convergence of F_n towards F. Similar proofs can be found in Chung (2001) and a similar result in Chow and Teicher (1988). *Q.E.D.*

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

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

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

$$\big[G\big(c_nS_n(\beta_0)\big) - \tilde{G}_n\big(c_nS_n(\beta_0)|X_n\big)\big] = o_p(1).$$

The same holds for G_n,

$$\big[G\big(c_nS_n(\beta_0)\big) - G_n\big(c_nS_n(\beta_0)|X_n\big)\big] = o_p(1).$$

Hence

$$G_n\big(c_nS_n(\beta_0)|X_n\big) = \tilde{G}_n\big(c_nS_n(\beta_0)|X_n\big) + o_p(1). \tag{A.26}$$

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

$$\mathsf{P}\big[G_n\big(c_nS_n(\beta_0)\big) \le \alpha\big] = \alpha. \tag{A.27}$$

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

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

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

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

as $\{\alpha_1 < G_n\big(c_nS_n(\beta_0)\big) \le \alpha\}$ is a zero-probability event. Applying (A.27) to α_1,

$$\mathsf{P}\big[G_n\big(c_nS_n(\beta_0)\big) \le \alpha\big] = \mathsf{P}\big[G_n\big(c_nS_n(\beta_0)\big) \le \alpha_1\big] = \alpha_1 \le \alpha. \tag{A.28}$$

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

$$\mathsf{P}\big[G_n\big(c_nS_n(\beta_0)\big) \le \alpha\big] \le \alpha. \tag{A.29}$$

Equation (A.29) combined with equation (A.26) allows us to write,

$$\mathsf{P}\big[\tilde{G}_n\big(c_nS_n(\beta_0)\big) \le \alpha\big] = \mathsf{P}\big[G_n\big(c_nS_n(\beta_0)\big) \le \alpha\big] + o_p(1) \le \alpha + o_p(1), \tag{A.30}$$

that is,

$$\lim_{n\to\infty} \mathsf{P}\big[\tilde{G}_n\big(c_nS_n(\beta_0)\big) \le \alpha\big] \le \alpha. \tag{A.31}$$

A.8. Proof of Theorem 6.4

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

$$\begin{aligned} \tilde{G}_n^N[x, n, S_n^{(0)}, S_n(N), U(N+1)|X_n] &= 1 - \frac{1}{N}\sum_{j=1}^{N} s(x - c_n S_n^{(j)}) \\ &+ \frac{1}{N}\sum_{j=1}^{N} \delta(c_n S_n^{(j)} - x) s(U^{(j)} - U^{(0)}), \end{aligned} \tag{A.32}$$

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

$$\tilde{p}_n^N(x) = \frac{N\tilde{G}_n^N(x) + 1}{N+1}. \tag{A.33}$$

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, we will show that 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 $\pi : [1, N] \rightarrow [1, N]$,

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

First, let rewrite

$$\mathsf{P}[S_n^{(0)} \geq t_0, S_n^{(1)} \geq t_1, \ldots, S_n^{(N)} \geq t_N] = \mathsf{E}_{X_n}\{\mathsf{P}[S_n^{(0)} \geq t_0, S_n^{(1)} \geq t_1, \ldots, S_n^{(N)} \geq t_N, X_n = x_n]\}.$$

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

$$\begin{aligned} \mathsf{P}[S_n^{(0)} \geq t_0, S_n^{(1)} \geq t_1, \ldots, S_n^{(N)} \geq t_N, X_n = x_n] &= \mathsf{P}[X_n = x_n]\prod_{i=0}^{N} \mathsf{P}[S_n^{(i)} \geq t_i | X_n = x_n] \\ &= G_n(t_0|X_n = x_n)\prod_{i=1}^{N} \tilde{G}_n(t_i|X_n = x_n). \end{aligned}$$

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

$$\mathsf{P}[S_n^{(0)} \geq t_0, S_n^{(1)} \geq t_1, \ldots, S_n^{(N)} \geq t_N, X_n = x_n] \rightarrow \prod_{i=0}^{N} G(t_i) \text{with probability one.} \tag{A.34}$$

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

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

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

$$\lim_{n\rightarrow\infty} \mathsf{P}[S_n^{(0)} \geq t_0, S_n^{(1)} \geq t_1, \ldots, S_n^{(N)} \geq t_N] - \mathsf{P}[S_n^{\pi(0)} \geq t_0, S_n^{\pi(1)} \geq t_1, \ldots, S_n^{\pi(N)} \geq t_N] = 0.$$

Hence, we can apply an asymptotic version of Proposition 2.2.2 in Dufour (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,

$$\lim_{n\rightarrow\infty} \tilde{p}_n^N(c_n S_n^{(0)}) \leq \alpha.$$

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

This appendix contains additional results for the Barro and Sala-i-Martin application. First, a residual analysis which includes outlier detection, heteroskedasticity tests, etc. is summarized in Table 8 and detailed in Table 9 and Figures 4-21. Second, complete sign-based inference results for the model parameters are reported in Tables 10 and 11.

Table 8. Regressions for personal income across U.S. States, 1880-1988: summary of regression diagnostics.

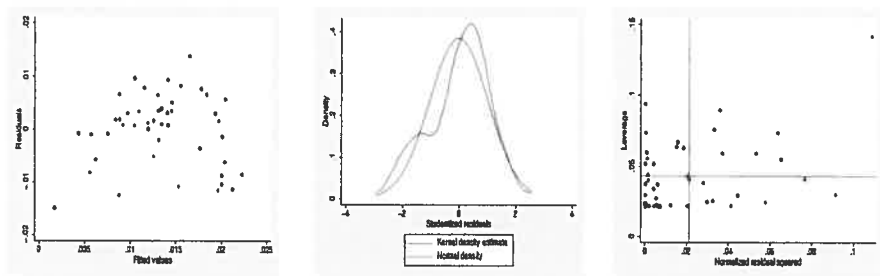
Period	Heterosked.*		Nonnormality**		Influent. obs.**		Possible outliers**	
	Basic eq.	*Eq Reg. Dum.*						
1880-1900	yes	-	yes	-	yes	yes	no	no
1900-1920	yes	yes	yes	yes	yes	yes	yes (MT)	yes
1920-1930	-	-	-	-	yes	-	no	no
1930-1940	-	-	yes	-	yes	yes	no	no
1940-1950	-	-	-	-	yes	yes	yes (VT)	yes (VT)
1950-1960	-	-	-	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 Breush-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, studendized or standardized residuals > 3, DF-beta and Cooks distance have been performed and lead to suspicions for nonnormality, outlier or high influential observation presence.

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	Breush-Pagan test	White test	Breush-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

Figure 4. Residual analysis: basic equation 1880-1900



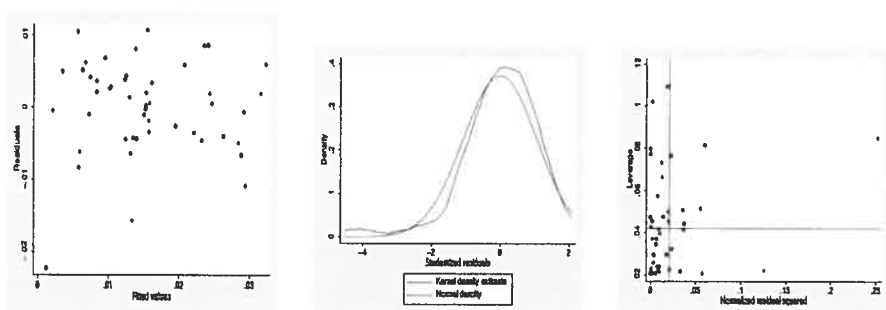
(a) Scatter plot (b) Kernel density (c) Leverage analysis

Skewness=-.506 Kurtosis=2.69

<u>Outliers detection results</u>: No.

Studentized residuals > 3	:	0
Standardized residuals > 3	:	0
DFbeta > 1	:	0
Cooks distance > .5	:	0

Figure 5. Residual analysis: basic equation 1900-1920



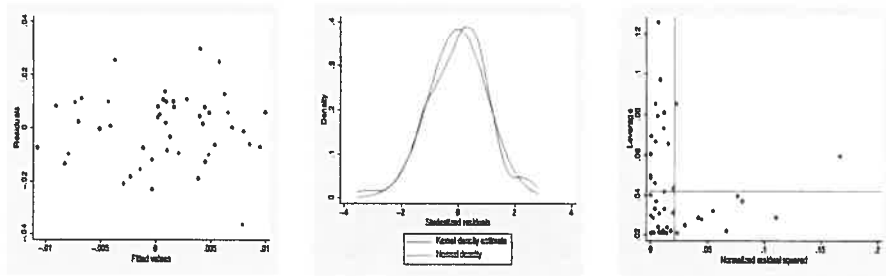
(a) Scatter plot (b) Kernel density (c) Leverage analysis

Skewness=-1.322 Kurtosis=6.23

<u>Outliers detection results</u>: yes : MT.

Studentized residuals > 3	:	MT
Standardized residuals > 3	:	MT
DFbeta > 1	:	MT
Cooks distance > .5	:	MT

Figure 6. Residual analysis: basic equation 1920-1930



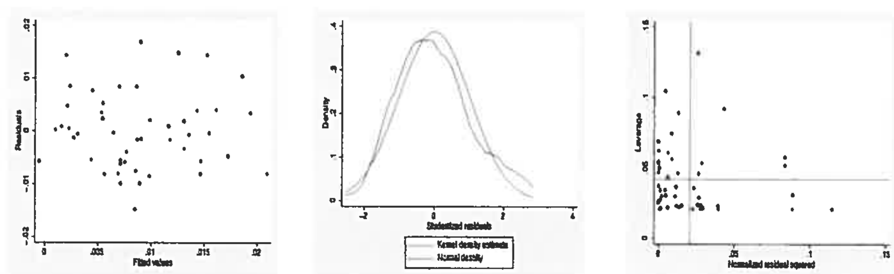
(a) Scatter plot (b) Kernel density (c) Leverage analysis

Skewness=-.309 Kurtosis=3.71

Outliers detection results: No.

Studentized residuals > 3	:	0
Standardized residuals > 3	:	0
DFbeta > 1	:	0
Cooks distance > .5	:	0

Figure 7. Residual analysis: basic equation 1930-1940



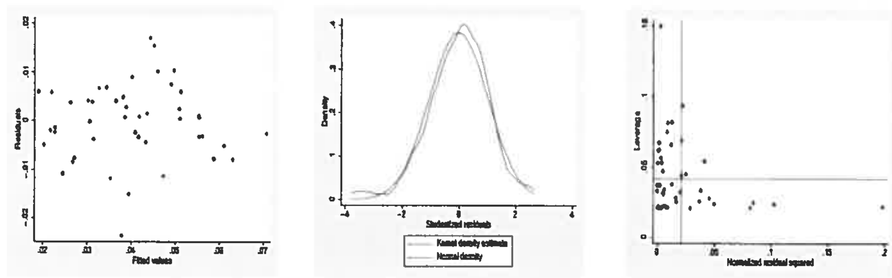
(a) Scatter plot (b) Kernel density (c) Leverage analysis

Skewness=-.495 Kurtosis=2.82

Outliers detection results: No.

Studentized residuals > 3	:	0
Standardized residuals > 3	:	0
DFbeta > 1	:	0
Cooks distance > .5	:	0

Figure 8. Residual analysis: basic equation 1940-1950



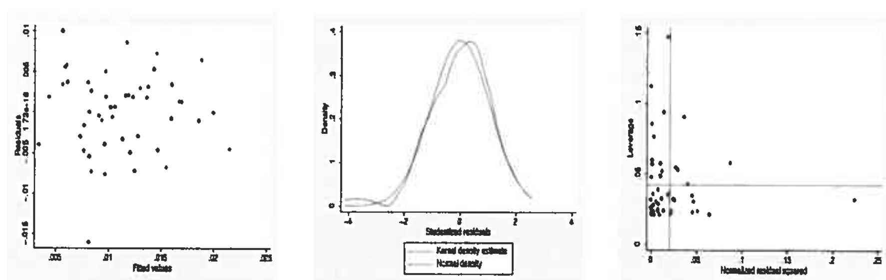
(a) Scatter plot (b) Kernel density (c) Leverage analysis

Skewness=-.550 Kurtosis=4.21

<u>Outliers detection results</u>: VT has a certain influence.

Studentized residuals > 3	:	VT
Standardized residuals > 3	:	VT
DFbeta > 1	:	0
Cooks distance > .5	:	0, but max for VT

Figure 9. Residual analysis: basic equation 1950-1960



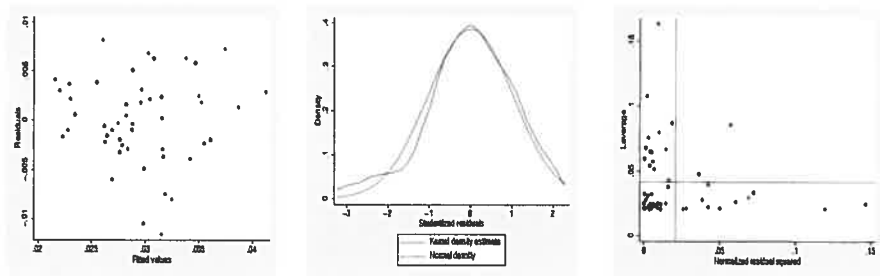
(a) Scatter plot (b) Kernel density (c) Leverage analysis

Skewness=-.745 Kurtosis=4.65

<u>Outliers detection results</u>: MT has a relative influence.

Studentized residuals > 3	:	MT
Standardized residuals > 3	:	MT
DFbeta > 1	:	0
Cooks distance > .5	:	0, but max for MT

Figure 10. Residual analysis: basic equation 1960-1970



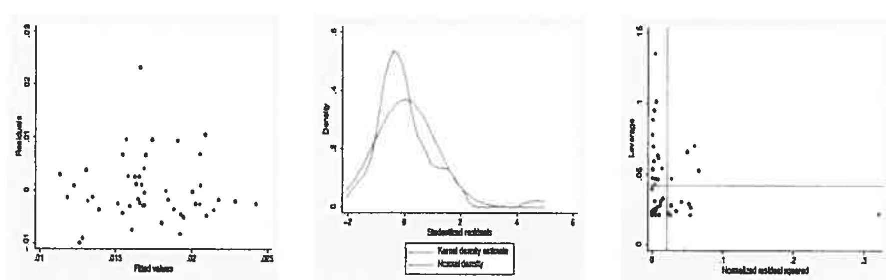
(a) Scatter plot (b) Kernel density (c) Leverage analysis

Skewness=-.518 Kurtosis=3.34

<u>Outliers detection results</u>: No.

Studentized residuals > 3	:	0
Standardized residuals > 3	:	0
DFbeta > 1	:	0
Cooks distance > .5	:	0

Figure 11. Residual analysis: basic equation 1970-1980



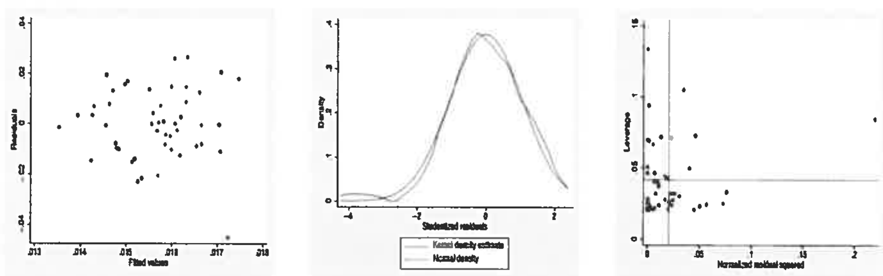
(a) Scatter plot (b) Kernel density (c) Leverage analysis

Skewness=1.790 Kurtosis=8.54

<u>Outliers detection results</u>: WY has some influence.

Studentized residuals > 3	:	WY
Standardized residuals > 3	:	WY
DFbeta > 1	:	0
Cooks distance > .5	:	0, but max for WY

Figure 12. Residual analysis: basic equation 1980-1988



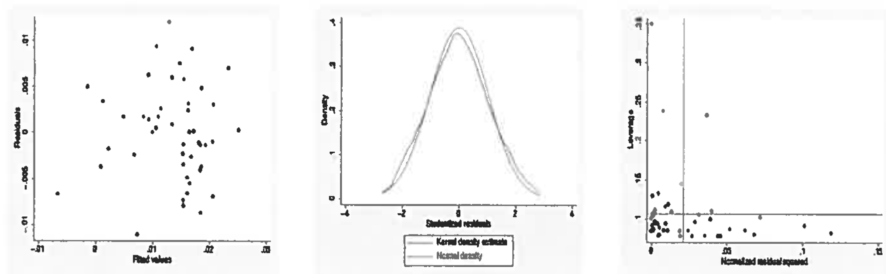
(a) Scatter plot (b) Kernel density (c) Leverage analysis

Skewness=-.707 Kurtosis=4.83

<u>Outliers detection results</u>: WY has a high influence.

Studentized residuals > 3	:	WY
Standardized residuals > 3	:	WY
DFbeta > 1	:	WY
Cooks distance > .5	:	WY

Figure 13. Residual analysis: regional dummies 1880-1900



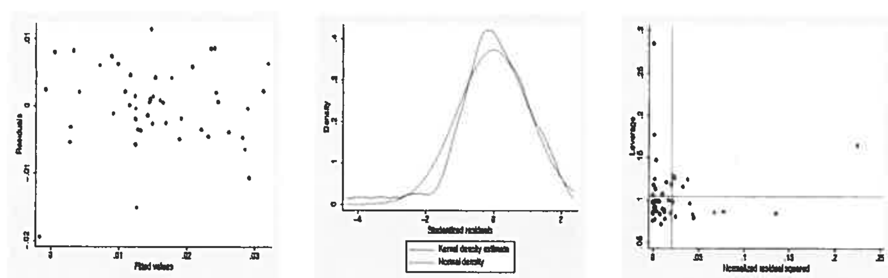
(a) Scatter plot (b) Kernel density (c) Leverage analysis

Skewness=.120 Kurtosis=2.72

Outliers detection results: No.

Studentized residuals > 3	:	0
Standardized residuals > 3	:	0
DFbeta > 1	:	0
Cooks distance $> .5$	:	0

Figure 14. Residual analysis: regional dummies 1900-1920



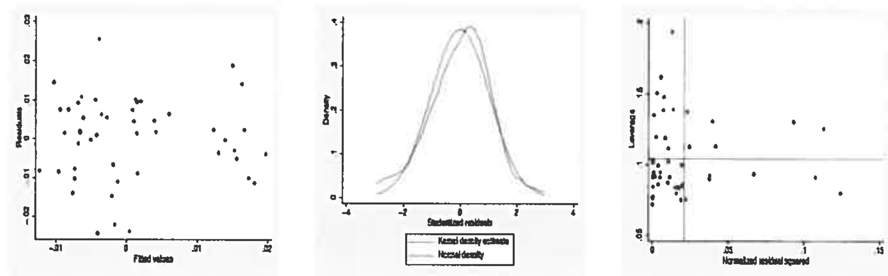
(a) Scatter plot (b) Kernel density (c) Leverage analysis

Skewness=-1.103 Kurtosis=5.73

Outliers detection results: yes : MT.

Studentized residuals > 3	:	MT
Standardized residuals > 3	:	MT
DFbeta > 1	:	MT on $DFbeta(y)$ and $DFbeta(south)$
Cooks distance $> .5$	:	MT

Figure 15. Residual analysis: regional dummies 1920-1930



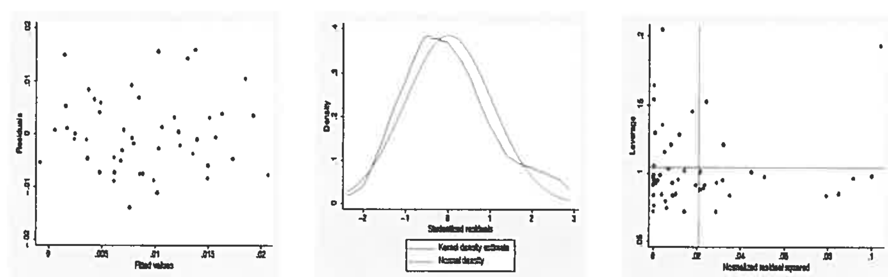
(a) Scatter plot (b) Kernel density (c) Leverage analysis

Skewness=-.316 Kurtosis=3.30

Outliers detection results: No.

Studentized residuals > 3	:	0
Standardized residuals > 3	:	0
DFbeta > 1	:	0
Cooks distance > .5	:	0

Figure 16. Residual analysis: regional dummies 1930-1940



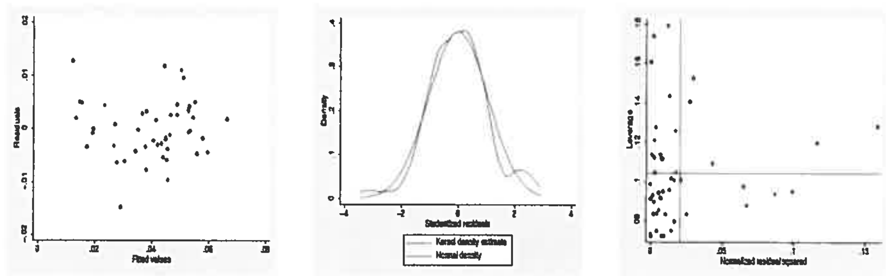
(a) Scatter plot (b) Kernel density (c) Leverage analysis

Skewness=.582 Kurtosis=2.92

Outliers detection results: No.

Studentized residuals > 3	:	0
Standardized residuals > 3	:	0
DFbeta > 1	:	0
Cooks distance > .5	:	0

Figure 17. Residual analysis: regional dummies 1940-1950



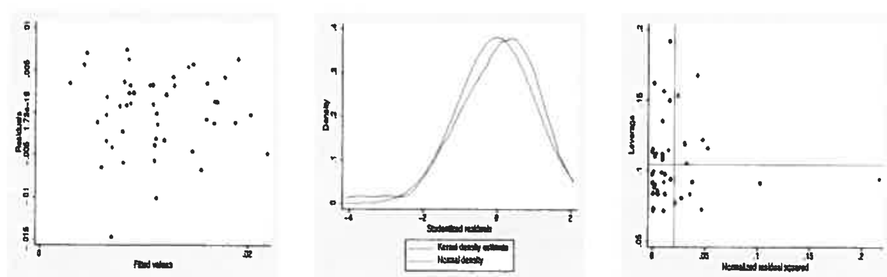
(a) Scatter plot (b) Kernel density (c) Leverage analysis

Skewness=.028 Kurtosis=3.90

<u>Outliers detection results</u>: VT has some influence.

Studentized residuals > 3	:	VT
Standardized residuals > 3	:	0
DFbeta > 1	:	0
Cooks distance > .5	:	0, but max for UT

Figure 18. Residual analysis: regional dummies 1950-1960



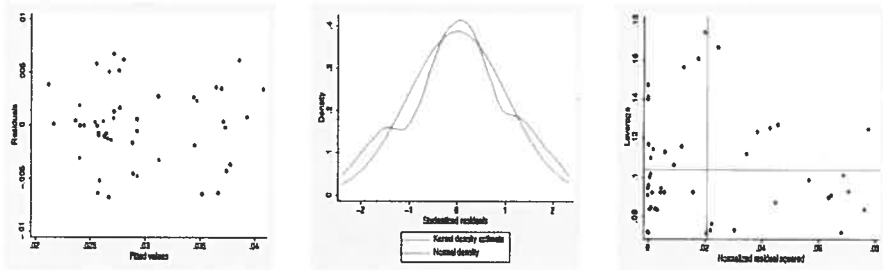
(a) Scatter plot (b) Kernel density (c) Leverage analysis

Skewness=-.975 Kurtosis=4.50

<u>Outliers detection results</u>: MT has some influence on y (and a high influence on dummies).

Studentized residuals > 3	:	MT
Standardized residuals > 3	:	MT
DFbeta > 1	:	MT with influence mostly is on a dummy
Cooks distance > .5	:	MT has influence

Figure 19. Residual analysis: regional dummies 1960-1970



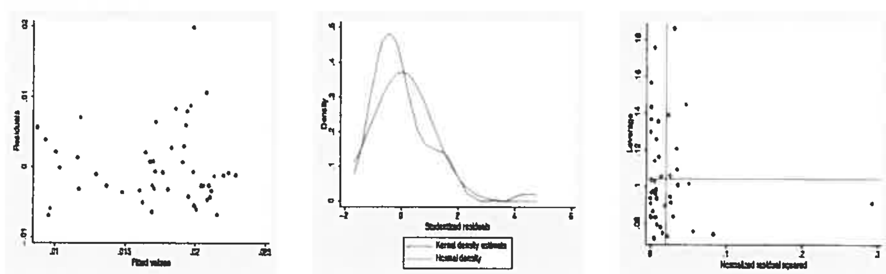
(a) Scatter plot (b) Kernel density (c) Leverage analysis

Skewness=-.093 Kurtosis=2.48

<u>Outliers detection results</u>: No.

Studentized residuals > 3	:	0
Standardized residuals > 3	:	0
DFbeta > 1	:	0
Cooks distance $> .5$	:	0

Figure 20. Residual analysis: Regional Dummies 1970-1980



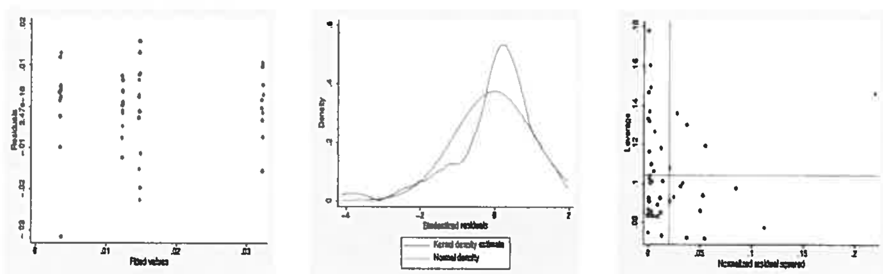
(a) Scatter plot (b) Kernel density (c) Leverage analysis

Skewness=1.708 Kurtosis=7.33

<u>Outliers detection results</u>: WY has a high influence on West but not on personal income.

Studentized residuals > 3	:	WY
Standardized residuals > 3	:	WY
DFbeta > 1	:	0, WY but on West
Cooks distance $> .5$	:	0, but max for WY

Figure 21. Residual analysis: regional dummies 1980-1988



(a) Scatter plot (b) Kernel density (c) Leverage analysis

Skewness=-1.25 Kurtosis=5.28

<u>Outliers detection results</u>: WY with high influence on personal income and West.

Studentized residuals > 3	:	WY
Standardized residuals > 3	:	WY
DFbeta > 1	:	WY nearly 1 for y and West
Cooks distance $> .5$	:	0, but max for WY

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

Table 11. Regressions for personal income across U.S. States, 1880-1988: complementary results.

Period	*Equation with regional dummies*		
	95% projection-based CI		
Variables:	*midwest*	*south*	*west*
1880-1900	[-.0091, -.0069]	[-.0109, -.0080]	[-.0110, -.0100]
1900-1920	[-.0130, -.0130]	[-.0248, -.0008]	[-.0218, -.0014]
1920-1930	[.0022, .0204]	[-.0038, .0404]	[-.0112, .0476]
1930-1940	[-.0074, -.0073]	[-.0345, .0105]	[-.0082, -.0010]
1940-1950	[-.0358, -.0385]	[-.0401, -.0124]	[-.0264, .0231]
1950-1960	[-.0187, -.0142]	[-.0283, -.0074]	[-.0152, -.0088]
1960-1970	[-.0178, -.0126]	[-.0319, -.0010]	[-.0194, .0177]
1970-1980	[-.0053, -.0015]	[-.0379, -.0045]	[-.0246, .0129]
1980-1988	[.0036, .0190]	[.0122, .0179]	[.0026, .0058]

C. Compared inference methods in simulations

Two *sign-based statistics* are studied: one adapted for mediangale process, $SF = D_S(\beta_0, (X'X)^{-1})$, see equation (C.35) and one corrected for serial dependence, $SHAC = D_S(\beta_0, \hat{J}_n^{-1})$, see equation (C.36).

$$D_S(\beta_0, (X'X)^{-1}) = s(y - X\beta_0)'X(X'X)^{-1}X's(y - X\beta_0). \tag{C.35}$$

$$D_S(\beta_0, \hat{J}_n^{-1}) = s(y - X\beta_0)'X\hat{J}_n^{-1}X's(y - X\beta_0). \tag{C.36}$$

where

$$\hat{J}_n = \frac{n}{n-p} \sum_{j=-n+1}^{n-1} k\left(\frac{j}{B_n}\right)\hat{\Gamma}_n(j), \tag{C.37}$$

with

$$\hat{\Gamma}_n(j) = \begin{cases} \frac{1}{n}\sum_{t=j+1}^{n} V_t(\beta_0)V'_{t-j}(\beta_0) & \text{for } j \geq 0 \\ \frac{1}{n}\sum_{t=-j+1}^{n} V_{t+j}(\beta_0)V'_t(\beta_0) & \text{for } j < 0, \end{cases} \tag{C.38}$$

and $V_t(\beta_0) = s(y_t - x_t'\beta_0) \times x_t$, $t = 1, \ldots, n$ and $k(.)$ is a real-valued kernel, here Bartlett kernel is used. The bandwidth parameter B_n is automatically adjusted [see 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 $H_0 : \beta = \beta_0$ are of the form

$$n(\hat{\beta} - \beta_0)\hat{D}_n^{-1}(\hat{\beta} - \beta_0) \tag{C.39}$$

where $\hat{D}_n$ is an estimate of the asymptotic covariance matrix for $\hat{\beta}$.
The OLS estimator is computed in GAUSS:

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'y. \tag{C.40}$$

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:

$$\hat{\beta}_{LAD} = \arg\min \sum_{t=1}^{n} |y_t - x_t'\beta|. \tag{C.41}$$

The following LAD covariance matrix estimators are considered.

The order statistic estimator (OS) [see Chamberlain (1994), Buchinsky (1995, 1998)] is valid for $i.i.d$ observations and is used as a benchmark. For $i.i.d$ observations, the LAD covariance matrix reduces to

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

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

$$\mathrm{P}[y_{(j)} \leq \xi_{1/2}] = \sum_{i=j}^{n} C_n^i (1/2)^n \tag{C.42}$$

entails

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

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,

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

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 allows one to estimate σ_{LAD}

$$\hat{\sigma}^2_{LAD} = \frac{n(Y_{int(n/2+l)} - Y_{int(n/2-l)})^2}{4Z^2_{1-\alpha/2}}.$$

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,

$$\hat{D}^{DMB} = \frac{m}{n}\left\{\frac{n}{B}\sum_{j=1}^{B}(\hat{\beta}^*_{LADj} - \hat{\beta}_{LAD})(\hat{\beta}^*_{LADj} - \hat{\beta}_{LAD})'\right\}. \qquad \text{(C.43)}$$

The moving block bootstrap centering around the sample estimate (MBB) was proposed by Fitzenberger (1997b). 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

$$\hat{D}^{MBB}(\hat{\beta}_{LAD}) = \frac{mb}{B}\left\{\sum_{j=1}^{B}(\hat{\beta}^*_{LADj} - \hat{\beta}_{LAD})(\hat{\beta}^*_{LADj} - \hat{\beta}_{LAD})'\right\}. \qquad \text{(C.44)}$$

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:

$$4\hat{f}_u(0)\left[\sum |y_i - x_i'\beta_0| - \sum |y_i - x_i'\hat{\beta}_{LAD}|\right] \tag{C.45}$$

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

Chapitre 2

Robust sign-based estimators and generalized confidence distributions in median regressions under heteroskedasticity and nonlinear dependence of unknown form

1. Introduction

The median regression framework is known to be more appropriate than the mean regression when unobserved heterogeneity or departure from normality is suspected in the data [see Dodge (1997)]. The associated estimators are more robust to outliers than usual least squares methods. They are also more efficient whenever the median is a better measure of location than the mean. This holds for heavy-tailed distributions or distributions possessing a mass at zero. The least absolute deviations (LAD) estimator has been widely studied in the literature and many papers have relaxed the distributional assumptions needed for consistency and asymptotic normality [see Powell (1984), Weiss (1991), Fitzenberger (1997b)]. The advantages of departing from a restrictive parametric framework is however reduced by the fact that inference is commonly based on asymptotic approximations (LAD asymptotic normality, Wald-type tests) in conjunction with kernel methods [Powell (1984)] or bootstrap procedures [design matrix bootstrap in Buchinsky (1995), block bootstrap in Fitzenberger (1997b), Bayesian bootstrap in Hahn (1997)].[1]

Asymptotic inference may be greatly misleading in small samples. Asymptotic tests may indeed present important size distortions. This point is well documented in LAD-based regressions [see Dielman and Pfaffenberger (1988a, 1988b), De Angelis, Hall, and Young (1993), Buchinsky (1995), Coudin and Dufour (2005a)] and time series [see Dufour (1981), Campbell and Dufour (1995, 1997)]. Asymptotic failures motivate us to adopt a different approach in the context of the median regression. In Coudin and Dufour (2005a), we focused on the testing problem. We developed a system of inference based on a general class of sign-based statistics, which allows one to conduct simultaneous tests on the complete vector of parameters with a fully controlled level for any sample size and under very weak distributional assumptions. Especially, the disturbance process may not be second-order stationary and may not possess a density. We assumed that the median of

[1]The reader is referred to Buchinsky (1995, 1998), for a review and to Fitzenberger (1997b) for a comparison between these methods. Other notable research on LAD estimators and their variants: the efficient weighted LAD of Zhao (2001), the smoothed LAD of Horowitz (1998), adaptations to allow for endogeneity [Amemiya (1982), Powell (1983), Hong and Tamer (2003)], nonlinear functional forms [Weiss (1991)] and generalization to quantile regressions [Koenker and Bassett (1978)]. Concerning empirical studies, Buchinsky (1994) used LAD and quantile estimators to study inequality and mobility in the U.S. labor market, and Engle and Manganelli (1999) provided an application in Value at Risk issues in finance. For reviews of the empirical literature on this topic, see Buchinsky (1998) and Koenker and Hallock (2000).

the current disturbance conditional on its own past and on the whole explanatory variable process is zero. The inference method is completely free of nuisance parameter, so Monte Carlo tests can be built.[2] This method does not require one to estimate the error density at zero in contrast to tests based on kernel estimates of the LAD asymptotic covariance matrix. Valid confidence regions and general tests are then derived by projection techniques [Dufour (1990, 1997), Dufour and Kiviet (1998), Abdelkhalek and Dufour (1998), Dufour and Jasiak (2001), Dufour and Taamouti (2005)]. Therefore, the test criteria are modified to cover linear dependence and the resulting inference is asymptotically valid.

The present paper introduces inference tools that can be associated with the previous system. First, the confidence distribution [Schweder and Hjort (2002)], which is a reinterpretation of Fisher fiducial distributions and the corresponding p-value function, yield the degree of confidence one may have in a certain value of the parameter. Second, the parameter value with the highest confidence (*i.e.* the highest p-value) provides a Hodges-Lehmann sign-based estimator [Hodges and Lehmann (1963)].

In frequentist econometrics, inference results are usually reported using confidence intervals and p-values [Neyman (1941)]. Fisher's fiducial distributions [Fisher (1930), Efron (1998)] are not commonly used. Fisher introduced the fiducial probability as a frequentist competitor to Bayesian posterior probabilities. Ignored for a long time, fiducial inference has recently enjoyed a renewed interest in the statistical literature with the introduction of confidence distributions and similar inference methods [see Hannig (2006) for a review]. The confidence distribution is defined in the one-dimensional model as a distribution whose quantiles span all the possible confidence intervals [Schweder and Hjort (2002)]. The latter authors introduced it as a Neymanian interpretation of Fisher's fiducial distribution. This tool summarizes all the inference results on the parameters and gives a graphical representation of them. Confidence distributions are not commonly used in econometrics for two reasons. First, they are only defined in the one-parameter case. Second, they require the exact distribution of the test statistic. The sign transform enables one to construct statistics that are pivots with known distribution without imposing parametric restrictions on the sample. Since the sign-based statistics are discrete, only approximate confidence distributions

[2] See Dwass (1957), Barnard (1963), Dufour (2006).

are obtained. The confidence distribution is related to the p-values for testing hypotheses of the form $H_0(\beta_0)$: $\beta = \beta_0$. The p-value can be seen as the *degree of confidence* one may have on the tested value. Our aim is to adapt those notions to a multidimensional parameter in a median regression context. For this, we shall combine sign-based tests of simultaneous hypothesis with increasing level with projection techniques. For each component, a projected p-value function provides a graphical illustration of both the inference summary and the degree of identification.

Then, we shall derive estimators and study their properties. Hodges and Lehmann (1963) proposed a general principle to directly derive estimators from test statistics for a given sample size.[3] They suggest to invert a test for $H_0(\beta_0)$: $\beta = \beta_0$, and to choose the value of β which is "least rejected" by the test. In a multidimensional context, this leads one to select the value of β with the highest *degree of confidence* $i.e.$ with the highest p-value. It is natural to associate the resulting sign-based estimator with a finite-sample-based inference method. This estimator also inherits some of the attractive properties of sign-based tests (robustness to model specification, gross errors and heteroskedasticity). We shall see that this estimator can be computed by minimizing quadratic forms of the constrained signs (with probability one). So it has a classical GMM form [Hansen (1982), and Honore and Hu (2004) for GMM statistics involving signs]. We show that sign-based estimators are consistent and asymptotically normal under regularity conditions weaker than the ones required by the LAD estimator usual theory. In particular, asymptotic normality and consistency hold for heavy-tailed disturbances that may not possess finite variance. This interesting property is entailed by the sign transformation. Signs of residuals always possess finite moments so no further restriction on the disturbance moments is required to complete the proofs. Contrary to usual GMM estimators, sign-based estimators are not just asymptotically justified by the analogy principle. They are first Hodges-Lehmann estimators associated with a finite-sample-based statistic.

The class of estimators so obtained includes some special cases studied in the statistical literature: Boldin, Simonova, and Tyurin (1997) derived *sign-estimators* from locally most

[3] First applied by Hodges and Lehmann to the Wilcoxon's signed rank-statistic for estimating a shift or a location, this principle was adapted for a regression context by Jureckova (1971), Jaeckel (1972) and Koul (1971). The latter authors derived so-called R-estimators from rank or signed-rank statistics.

powerful test statistics for $i.i.d.$ observations and fixed regressors. In such a context, they proved consistency and asymptotic normality. More precisely, assuming $i.i.d.$ observations and fixed regressors, they showed a well standardized sign-estimator is asymptotically equivalent to the LAD estimator. However, in finite samples and for various setups, LAD and sign methods exhibit very different features. The simulation studies of bias and root mean squared error (RMSE) we present show that sign-based estimators are more robust than the LAD estimator in the presence of heteroskedasticity. We shall therefore compare both their finite-sample and asymptotic properties. We will provide consistency when point identification is available, asymptotic normality and a Monte Carlo study of performance.

Instrumental versions of sign-based estimators are presented in Honore and Hu (2004) and Hong and Tamer (2003). Honore and Hu (2004) derived the so-called *median-based estimator* as an instrumental GMM version of the quantile estimator. The authors motivated to use the latter along with other rank-based estimators for their general robustness properties. However, the major advantage of signs upon ranks is to easily deal with heteroskedastic disturbances. In the present paper, we do not assume $i.i.d.$ disturbances. We derive various sign-based statistics and associated sign-based estimators depending on the setup. Many heteroskedastic and possibly dependent schemes are covered and, when needed, an heteroskedasticity and autocorrelation correction is included in the estimator criterion function. Restricting on $i.i.d.$ cases, Honore and Hu (2004) observed in simulations that inference based on rank-based estimators performed better than the median-based one. In particular, the estimates of the asymptotic standard errors of the median-based estimator, that they obtained by kernel, were too small and the associated inference suffered from overrejection of the null hypothesis. Deriving sign-based estimators as Hodges-Lehmann estimators motivates us to definitely combine them to the inference method they come from. The latter, based on the exact distribution of the corresponding sign-based test statistics does not depend on any nuisance parameter and does control test levels [see Coudin and Dufour (2005a)]. Finally, sign-based tests, projection-based confidence regions, projection-based p-values and sign-based estimators constitute a whole system of inference valid for any given sample size under very weak distributional assumptions.

The paper is organized as follows. Section 2 presents the model, the sign-based statistics and the Monte Carlo tests. Section 3 is dedicated to confidence distributions and p-value functions. In section 4, we introduce the sign-based estimators, which are obtained by maximizing the p-value function. Finite-sample and asymptotic properties of sign-based estimators are established in section 5. In section 6, we present a simulation study of bias and RMSE. In section 7, we apply sign-based estimation for deriving robust estimates in two cases: first, in a financial setup involving large heteroskedasticity (S.&P. index); second, in a cross-sectional regional data set where the sample size is necessarily small (β-convergence of output levels across U.S. States). Section 8 concludes. Appendix A contains the proofs.

2. Framework

2.1. Model

We consider the framework of Coudin and Dufour (2005a). Let $\{W_t = (y_t,\ x_t') : \Omega \to \mathbb{R}^{p+1}\}_{t=1,2,\dots}$ be a stochastic process defined on a probability space $(\Omega, \mathcal{F}, P)$. $\{W_t, \mathcal{F}_t\}_{t=1,2,\dots}$ is an adapted stochastic sequence, *i.e.*, $\mathcal{F}_t$ is a σ-field in Ω such that $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s < t$ and $\sigma(W_1, \dots, W_t) \subset \mathcal{F}_t$, where $\sigma(W_1, \dots, W_t)$ is the σ-algebra spanned by $W_1, \dots, W_t$. We set $W_t = (y_t, x_t')$, where y_t is the dependent variable and $x_t = (x_{t1}, \dots, x_{tp})'$, a p-vector of explanatory variables. The x_t's may be random or fixed. We assume that y_t and x_t satisfy a simple linear model of the form:

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

or, in vector notation,

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

where $y = (y_1, \dots, y_n)'$ and $u = (u_1, \dots, u_n)'$ are $n \times 1$ real vectors, $X = [x_1, \dots, x_n]'$ is an $n \times p$ real matrix. $\beta \in \mathbb{R}^p$ is the vector of parameters. The u_t's can be heteroskedastic each one with conditional distribution function denoted $F_t(.|X)$:

$$u_t|X \sim F_t(.|X), \quad t = 1, \dots, n.$$

The traditional form of a median regression assumes u_t's are $i.i.d.$ with median zero

$$\text{Med}(u_t|x_1,\ldots,x_n) = 0, \;\; t = 1,\ldots,n. \tag{2.3}$$

Here, we relax the assumption that the u_t are $i.i.d.$ and consider instead moment conditions based on residual signs where the sign operator $s : \mathbb{R} \to \{-1, 0, 1\}$ defined as

$$s(a) = \mathbf{1}_{[0,+\infty)}(a) - \mathbf{1}_{(-\infty,0]}(a), \;\; a \in \mathbb{R}. \tag{2.4}$$

For convenience, the notation will be extended to vectors. Let $u \in \mathbb{R}^n$ and $s(u)$, the n-vector composed by the signs of its components. We assume the following assumption holds.

Assumption A1 SIGN MOMENT CONDITION. $E[s(u_t)x_{kt}] = 0,$ *for* $k = 1,\ldots,p,\; t = 1,\ldots,n,$ *and* $n \in \mathbb{N}$.

Assumption A1 is fulfilled if the disturbances are $i.i.d.$ and more generally if the signs satisfy a martingale difference with respect to the past information $\mathcal{F}_t = \sigma(W_1,\ldots,W_t)$:

$$E[s(u_t)|\mathcal{F}_{t-1}] = 0, \;\; \forall t \geq 1. \tag{2.5}$$

Assumption A1 also covers many weakly dependent processes including usual linear dependent processes, such as $AR(1)$ disturbances with normal innovations and mean zero. This has been pointed out by Fitzenberger (1997b). Next, Assumption A1 holds when u satisfies the strict conditional mediangale condition defined in Coudin and Dufour (2005a):

Assumption A2 STRICT CONDITIONAL MEDIANGALE. *Let* $\{u_t, \mathcal{F}_t\}_{t=1,\ldots}$ *be an adapted stochastic sequence where* $\mathcal{F}_t = \sigma(u_1,\ldots,u_t,X)$. $\{u_t\}_{t=1,\ldots}$ *is a strict mediangale conditional on* X *with respect to* $\{\mathcal{F}_t\}_{t=1,\ldots}$ *if*

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

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

Assumption A2 and more generally the moment condition A1 are exploited to construct test statistics.

2.2. Sign-based statistics and Monte Carlo tests

For testing $H_0(\beta_0)$: $\beta = \beta_0$ vs. $H_1(\beta_0)$: $\beta \neq \beta_0$ in model (2.1), we consider general quadratic forms involving the vector of the residual signs for the constrained model $s(y - X\beta_0)$:

$$D_S(\beta_0, \Omega_n) = s(y - X\beta_0)'X\Omega_n(s(y - X\beta_0), X)X's(y - X\beta_0) \tag{2.6}$$

where $\Omega_n(s(y - X\beta_0), X)$ is a $p \times p$ positive definite weight matrix that may depend on the constrained signs. In Coudin and Dufour (2005a), we developed distribution-free Monte Carlo tests under the mediangale Assumption A2. We briefly summarize it.

If the disturbances satisfy the mediangale Assumption A2, the sign-based statistics satisfying equation (2.6) are shown to be pivotal functions under $H_0(\beta_0)$. The distribution of the statistic conditional on the realization of X, is perfectly specified and can be simulated. Monte Carlo tests with controlled levels are constructed in the following way. For testing $H_0(\beta_0)$: $\beta = \beta_0$ vs. $H_1(\beta_0)$: $\beta \neq \beta_0$ with level $\alpha \in [0, 1]$, we denote $D_S^{(0)} = D_S(\beta_0)$ the observed statistics, $(D_S^{(1)}, \ldots, D_S^{(N)})'$ an N-vector of independent replicates drawn from the same distribution as $D_S(\beta_0)$ and $(W^{(0)}, \ldots, W^{(N)})'$, a $N + 1$-vector of $i.i.d.$ uniform variables. A Monte Carlo test for $H_0(\beta_0)$ consists in rejecting the null hypothesis whenever the empirical p-value $\tilde{p}_N^{D_S}(\beta_0)$ is smaller than α, where

$$\tilde{p}_N^{D_S}(\beta_0) = 1 - \frac{1}{N+1}\left[\sum_{j=1}^{N} \mathbf{1}_{[0,\,\infty)}(D_S^{(0)} - D_S^{(j)}) - \sum_{j=1}^{N} \mathbf{1}_{\{0\}}(D_S^{(j)} - D_S^{(0)})\mathbf{1}_{[0,\,\infty)}(W^{(j)} - W^{(0)})\right]. \tag{2.7}$$

This empirical p-value is well adapted to discrete statistics. When two realizations of the statistic are equal, they are ordered using the auxiliary continuous uniform variables $W^{(j)}$ (randomized tie-breaking). When the number of replicates N is such that $\alpha(N + 1)$ is an integer, the level of the Monte Carlo test is equal to α for any sample size n [see Dufour (2006), Coudin and Dufour (2005a)]. Next, simultaneous confidence regions for the entire parameter β are obtained by inverting those simultaneous tests. The simultaneous confidence region $C_{1-\alpha}(\beta)$ with level $1 - \alpha$ contains all the values β^* with empirical p-value

$\tilde{p}_N^{D_S}(\beta^*)$ [associated with the test of $H_0(\beta^*): \ \beta = \beta^*$] higher than α:

$$C_{1-\alpha}(\beta) = \{\beta^* | \tilde{p}_N^{D_S}(\beta^*) \geq \alpha\}.$$

By construction, this confidence region has level $1 - \alpha$ for any sample size. It is then possible to derive general (and possibly nonlinear) tests and confidence sets by projection techniques. For example, individual confidence intervals are obtained in such a way. If D_S is an asymptotically pivotal function all previous results hold asymptotically. For a detailed presentation, see Coudin and Dufour (2005a).

3. Confidence distributions

In the one parameter model, statisticians have defined the confidence distribution notion that summarizes a family of confidence intervals; see Schweder and Hjort (2002). By definition, the quantiles of a confidence distribution span all the possible confidence intervals of a real β. The confidence distribution is a reinterpretation of the Fisher fiducial distributions and provides, in a sense, an equivalent to Bayesian posterior probabilities in a frequentist setup [see also Fisher (1930), Neyman (1941) and Efron (1998)]. This statistical notion is not commonly used in the econometric literature, for two reasons. First, it is only defined in the one-parameter case. Second, it requires that the test statistic be a pivot with known exact distribution. Our aim is to extend that notion (or an equivalent) to multidimensional parameters. The sign transformation enables one to construct statistics which are pivots with known distribution without imposing parametric restrictions on the sample. Consequently, our setup does not suffer from the second restriction. In that section, we briefly recall the initial statistical concept and apply it to an example in univariate regression. Then, we address the extension to multidimensional regressions.

3.1. Confidence distributions in univariate regressions

Schweder and Hjort (2002) defined the confidence distribution for the real parameter β such a distribution depending on the observations (y, x), whose cumulative distribution function evaluated at the true value of β has a uniform distribution whatever the true value of β. In a formalized way, this can be expressed as follows:

Definition 3.1 CONFIDENCE DISTRIBUTION. *Any distribution with cumulative $CD(\beta)$ and quantile function $CD^{-1}(\beta)$, such that*

$$P_\beta[\beta \le CD^{-1}(\alpha; y; x)] = P_\beta[CD(\beta; y; x) \le \alpha] = \alpha \tag{3.1}$$

for all $\alpha \in (0,1)$ and for all probability distributions in the statistical model, is called a confidence distribution of β.

$(-\infty,\ CD^{-1}(\alpha)]$ constitutes a one-sided stochastic confidence interval with coverage probability α,[4] and the realized confidence $CD(\beta_0; y; x)$ is the p-value of the one-sided hypothesis $H_0:\ \beta \le \beta_0$ versus $H_1:\ \beta > \beta_0$ when the observed data are $y,\ x$. The realized p-value when testing $H_0:\ \beta = \beta_0$ versus $H_1:\ \beta \ne \beta_0$ is $2\min\{CD(\beta_0), 1 - CD(\beta_0)\}$. Those relations are stated in Lemma 2 of Schweder and Hjort (2002), which states: *the confidence of the statement "$\beta \le \beta_0$" is the degree of confidence $CD(\beta_0)$ for the confidence interval $\big(-\infty, CD^{-1}(CD(\beta_0))\big]$, and is equal to the p-value of a test of $H_0 : \beta \le \beta_0$ v.s. $H_1 : \beta > \beta_0$.* Hence, tests and confidence intervals on β are contained in the confidence distribution. Moreover, the values associated with the highest confidence statement (or equivalently with the highest p-value for testing $H_0 : \beta = \beta_0$) may provide estimators of β.

Schweder and Hjort (2002) also note that, since the cumulative function $CD(\beta)$ is an invertible function of β and is uniformly distributed, $CD(\beta)$ constitutes a pivot conditional on x. Reciprocally, whenever a pivot increases with β (for example a continuous statistic $T(\beta)$ with cumulative distribution function F that is independent of β and free of any nuisance parameter), $F\big(T(\beta)\big)$ is uniformly distributed and satisfies conditions for providing a confidence distribution. Let $\hat{\beta}$ be such a continuous real statistic increasing with β with a free of nuisance parameter distribution. A test of $H_0 : \beta \le \beta_0$ is rejected when $\hat{\beta}^{obs}$ is

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

large, with p-value $P_{\beta_0}[\beta > \hat{\beta}^{obs}]$. Then,

$$P_{\beta_0}[\beta > \hat{\beta}^{obs}] = 1 - F_{\beta_0}(\hat{\beta}^{obs}) = CD(\beta_0) \tag{3.2}$$

where $F_{\beta_0}(\hat{\beta})$ is the sampling distribution of $\hat{\beta}$. Consequently, simulated sampling distributions and simulated realized p-values as presented previously yield a way to construct simulated confidence distributions. The sampling distribution and the confidence distribution are fundamentally different theoretical notions. The sampling distribution is the probability distribution of $\hat{\beta}$ obtained by repeated samplings whereas the confidence distribution is an ex-post object which contains the confidence statements one can have on the value of β given $y, x, \hat{\beta}^{obs}$.

A last remark relates to discrete statistics. Confidence distributions based on discrete statistics cannot lead to a continuous uniform distribution. Approximations must be used. Schweder and Hjort (2002) proposed half correction. For discrete statistics, they used

$$CD(\beta_0) = P_{\beta_0}[\beta > \hat{\beta}^{obs}] + \frac{1}{2}P_{\beta_0}[\beta = \hat{\beta}^{obs}], \tag{3.3}$$

We rather use randomization as in section 2. The discrete statistic $\hat{\beta}$ is associated with an auxiliary one $U_{\hat{\beta}}$, which is independently, uniformly and continuously distributed over $[0, 1]$. Lexicographical order is used to order ties.

$$CD(\beta_0) = P_{\beta_0}[\beta > \hat{\beta}^{obs}] + P[U_{\hat{\beta}}^{(0)} > U_\beta]P_{\beta_0}[\beta = \hat{\beta}^{obs}], \tag{3.4}$$

Let us consider a simple example to illustrate those notions. In the model $y_i = \beta x_i + u_i,\ i = 1, \ldots, n$, $(u_i, x_i) \overset{iid}{\sim} \mathcal{N}(0, I_2)$, the Student sign-based statistic

$$\widehat{SST(\beta)} = \frac{\sum s(y_i - x_i\beta)x_i}{(\sum x_i^2)^{1/2}}$$

is a pivotal function and decreases with β. The confidence distribution of β given the realization y, x can be approximated by

$$\widehat{CD}(\beta_0) = 1 - \hat{F}_{\beta_0}(\widehat{SST(\beta_0)}), \tag{3.5}$$

with $\hat{F}_{\beta_0}$ a Monte Carlo estimate of the sampling distribution of SST under $\beta = \beta_0$. Figure 22 presents a simulated confidence distribution cumulative function for β, given

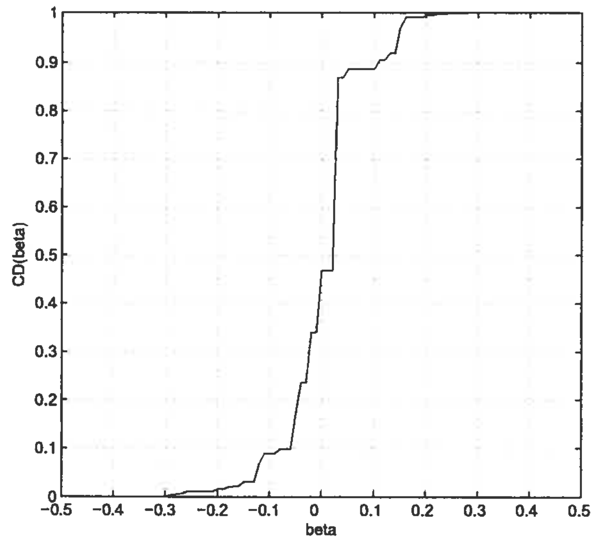


Figure 22. Simulated confidence distribution cumulative function based on SST.

200 realizations of (u_i, x_i) based on SST. The Monte Carlo estimate of $\hat{F}_{\beta_0}$ is obtained from 9999 replicates of SST under H_0. Testing $H_0 : \beta \leq .1$ at 10% can be done by reading $CD(.1)$, which equals the p-value of H_0, here .88. The test accepts H_0. Further, $(-\infty, .15]$ constitutes a one-sided confidence interval for β with level .95.

Another interesting object is the realized p-value function when testing point hypotheses $H_0(\beta_0) : \beta = \beta_0$. The latter is a simple transformation of the CD cumulative function:

$$\hat{p}_{SST}(\beta_0) = 2\min\{\widehat{CD}_{SST}(\beta_0), 1 - \widehat{CD}_{SST}(\beta_0)\}. \tag{3.6}$$

Consider now the statistic $SF = SST^2$. SF is a pivotal function but not a monotone function of β contrary to SST. An entire confidence distribution cannot be recovered from SF because of this lack of monotonicity. However, the p-value function can be constructed using equation (2.7). Figure 23 compares p-value functions based on SST and SF. Inverting the p-value function allows one to recover half of the confidence distribution and consequently half of the inference results, *i.e.* the two-sided confidence intervals. For example, $[-.12, .14]$ constitutes a confidence interval with level 90% for both statistics. The p-value

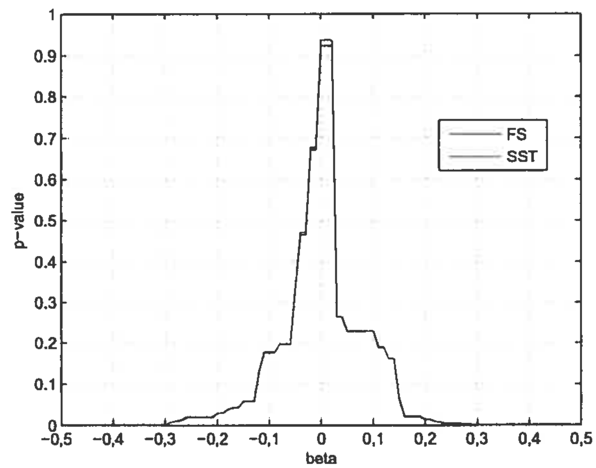


Figure 23. Simulated p-value functions based on SST and SF.

function provides then an interesting summary on the available inference. Especially, it gives the confidence degree one can have in the statement $\beta = \beta_0$.

The spread of the p-value function is also related to the *parameter identification*. When the p-values are low (or high) whatever the value of β, one may expect the parameter to be badly identified either because there exists a set of observationally equivalent parameters, then, the p-values are high for a wide set of values; either because there does not exist any value satisfying the model and then the p-values are small everywhere. To illustrate that point, let us consider an example where the first n_1 observations satisfy $y_i = \beta_1 x_i + u_i,\ i = 1, \ldots, n_1, (u_i, x_i) \overset{iid}{\sim} \mathcal{N}(0, I_2)$ and the n_2 followings, $y_i = \beta_2 x_i + u_i,\ i = n_1 + 1, \ldots, n_1 + n_2, (u_i, x_i) \overset{iid}{\sim} \mathcal{N}(0, I_2)$, with $\beta_1 = -1$ and $\beta_2 = 1$. The model $y_i = \beta x_i + u_i,\ i = 1, \ldots, n_1 + n_2$, is misspecified. In Figure 24, we notice the spread of the p-value function based on SF is large which we can interpret as a lack of identification: the set of observationally equivalent β is not reduced to a point.

The p-value function has an important advantage over the confidence distribution: it is straightforwardly extendable to multidimensional parameters.

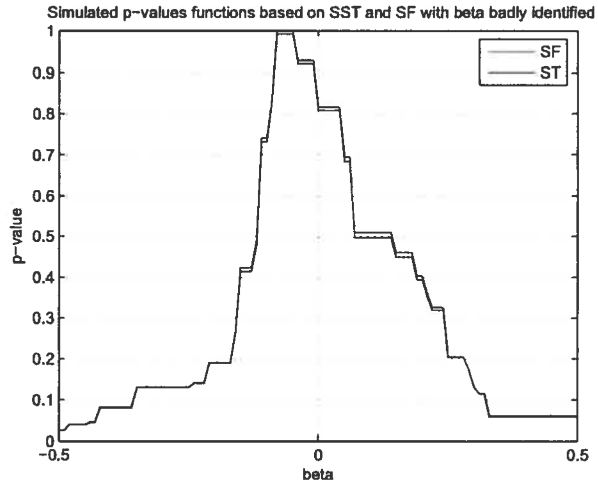


Figure 24. Simulated p-value functions based on SST and SF when the parameter is badly identified.

3.2. Simultaneous and projection-based p-value functions in multivariate regressions

If $p \geq 2$, the confidence distribution notion is not defined anymore. However, simulated p-values for testing $H_0 : \beta = \beta_0$ can easily be constructed from the SF statistic and more generally from any sign-based statistic which satisfies equation (2.6). Simulated p-values lead to a mapping for which we have a 3-dimensional representation for $p = 2$. Consider the model: $y_i = \beta^1 x_{1i} + \beta^2 x_{2i} + u_i,\ i = 1, \dots, n, (u_i, x_{1i}, x_{2i}) \overset{iid}{\sim} \mathcal{N}(0, I_3)$, $\beta = (0, 0)'$, $y = (y_1, \dots, y_n)'$, $u = (u_1, \dots, u_n)'$, $x_1 = (x_{11}, \dots, x_{1n})'$, $x_2 = (x_{21}, \dots, x_{2n})'$ and $X = (x_1, x_2)$. Let $D_S(\beta, (X'X)^{-1}) = s'(y - X\beta)X(X'X)^{-1}X's(y - X\beta)$. In Figure 25, we compute the simulated p-value function $\tilde{p}_N^{D_S}(\beta_0)$ for testing $H_0 : \beta = \beta_0$ on a grid of values of β_0, using N replicates of the sign vector. $\tilde{p}_N^{D_S}(\beta_0)$ allows one to construct simultaneous confidence sets for $\beta = (\beta^1, \beta^2)$ with any level. By construction, the confidence region $C_{1-\alpha}(\beta)$ defined as

$$C_{1-\alpha}(\beta) = \{\beta | \tilde{p}_N^{D_S}(\beta_0) \geq \alpha\}, \tag{3.7}$$

has level $1 - \alpha$ [see Dufour (2006)]. Hence, by construction, $C_{1-\alpha}(\beta)$ corresponds to the intersection of the horizontal plan at ordinate α with the envelope of $\tilde{p}_N^{D_S}(\beta_0)$. For higher

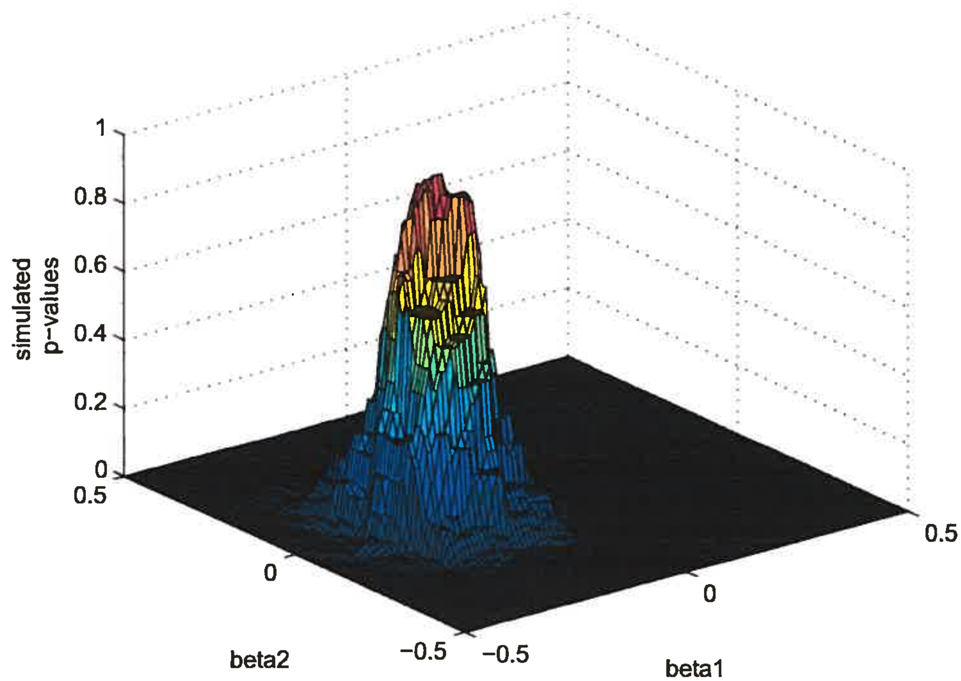
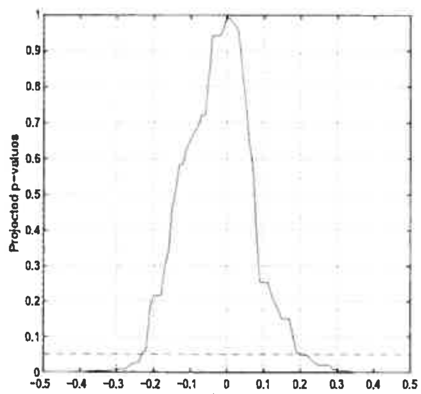


Figure 25. Simulated p-value functions based on SF ($n = 200$, $N = 9999$).

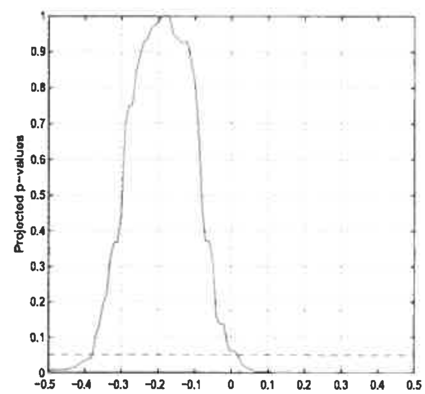
dimensions ($p > 2$), a complete graphical representation is not available anymore. However, one can consider projection-based p-value functions for each individual component of the parameter of interest in a similar way than projection-based confidence intervals. For this, we apply the general strategy of projection on the complete simultaneous p-value function. The projected-based p-value function for the component β^1 is given by:

$$\mathrm{P}.\tilde{\mathrm{p}}^{\beta^1}(\beta_0^1) = \max_{\beta_0^2 \in \mathbb{R}} \tilde{p}_N^{D_S}[(\beta_0^1, \beta_0^2)]. \tag{3.8}$$

Figure 26 presents projection-based confidence intervals for the individual parameters of the previous 2-dimensional example. $[-.22, .21]$ is a 95% (conservative) confidence interval for β^1. $[-.38, .02]$ is a 95% (conservative) confidence interval for β^2. Testing $\beta^1 = 0$ is accepted at 5% with p-value 1.0. Testing $\beta^2 = 0$ is accepted at 5% with p-value .06.



(a) Projection-based $p$values for β^1



(b) Projection-based $p$values for β^2

Figure 26. Projection-based p-values.

4. Sign-estimators

Sign-based estimators complete the above system of inference. Intuition suggests to consider values with the highest confidence degree, *i.e*, with the highest p-values. Estimators obtained by that sort of test inversion constitute multidimensional extensions of the Hodges-Lehmann principle.

4.1. Sign-based estimators as maxima of the p-value function

Hodges and Lehmann (1963) presented a general principle to derive estimators by test inversion; see also Johnson, Kotz, and Read (1983). Suppose $\mu \in \mathbb{R}$ and $T(\mu_0, W)$ is a statistic for testing $\mu = \mu_0$ against $\mu > \mu_0$ based on the observations W. Suppose further that $T(\mu, W)$ is nondecreasing in the scalar μ. Given a known central value of $T(\mu_0, W)$, say $m(\mu_0)$ [for example $E_W T(\mu_0, W)$], the test rejects $\mu = \mu_0$ whenever the observed T is larger than, say, $m(\mu_0)$. If that is the case, one is inclined to prefer higher values of μ. The reverse holds when testing the opposite. If $m(\mu_0)$ does not depend on μ_0 [$m(\mu_0) = m_0$], an intuitive estimator of μ (if it exists) is given by μ^* such that $T(\mu^*, W)$ equals m_0 (or is very close to m_0). μ^* may be seen as the value of μ which is most supported by the observations.

This principle can be directly extended to multidimensional parameter setups through p-value functions. Let $\beta \in \mathbb{R}^p$. Consider testing $H_0 : \beta = \beta_0$ versus $H_1 : \beta = \beta_1$ with the positive statistic T. A test based on T rejects H_0 when $T(\beta_0)$ is larger than a certain critical value that depends on the test level. The estimator of β is chosen as the value of β least rejected when the level α of the test increases. This corresponds to the highest p-value. If the associated p-value for $\beta = \beta_0$ is $\mathrm{p}(\beta_0) = G\big(D_S(\beta_0)|\beta_0\big)$, where $G(x|\beta_0)$ is the survival function of $D_S(\beta_0)$, *i.e.* $G(x|\beta_0) = \mathrm{P}[D_S(\beta_0) > x]$, the set

$$M1 = \underset{\beta \in \mathbb{R}^p}{\arg\max}\ \mathrm{p}(\beta) \tag{4.1}$$

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

4.2. Sign-based estimators as solutions of optimization problems

When the distribution of $T(\beta_0)$ and the corresponding p-value function do not depend on the tested value β_0, maximizing the p-value is equivalent to minimizing the statistic $T(\beta_0)$. This point is stated in the following proposition. Let us denote $\bar{F}(x|\beta_0)$ the distribution of $T(\beta_0)$ when $\beta = \beta_0$ and assume this distribution is invariant to β (Assumption A3).

Assumption A3 INVARIANCE OF THE DISTRIBUTION FUNCTION.

$$\bar{F}(x|\beta_0) = \bar{F}(x) \quad \forall x \in \mathbb{R}^+,\ \forall \beta_0 \in \mathbb{R}^p.$$

Then, the following proposition holds.

Proposition 4.1 *Under Assumption A3, the two following sets $M1$ and $M2$ are equal with probability one:*

$$M1 = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmax}}\ \mathrm{p}(\beta). \tag{4.2}$$

$$M2 = \underset{\beta \in \mathbb{R}^p}{\arg\min}\ T(\beta). \tag{4.3}$$

Maximizing $\mathrm{p}(\beta)$ is equivalent (in probability) to minimizing $T(\beta)$ if Assumption A3 holds. Under the mediangale Assumption A2, any sign-based statistic D_S does satisfy Assumption A3. Consequently,

$$\hat{\beta}_n(\Omega_n) \in \underset{\beta \in \mathbb{R}^p}{\arg\min}\ s'(Y - X\beta)X\Omega_n(s(Y - X\beta), X)X's(Y - X\beta) \tag{4.4}$$

equals (with probability one) a Hodges-Lehmann estimator based on $D_S(\Omega_n, \beta)$. Since $D_S(\Omega_n, \beta)$ is non-negative, problem (4.4) always possesses at least one solution. As signs can only take 3 values, for fixed n, the quadratic function can take a finite number of values, which entails the existence of the minimum. If the solution is not unique, one may add a choice criterion. For example, one can choose the smallest solution in terms of a norm or use a randomization. Under conditions of point identification, any solution of (4.4) is a consistent estimator.

The whole argmin set of (4.4) remains informative in models with sets of observationally equivalent values of β [see Chernozhukov, Tamer, and Hong (2006)]. The identified feature of those models is a set instead of a point value. Any inference approach relying on the consistency of a point estimator (which assumes point identification), gives misleading results, but the estimation of the whole set can be exploited. Let us remind that the Monte Carlo sign-based inference method [Coudin and Dufour (2005a)] does not rely on identification conditions and leads to valid results in any case.

The sign-based estimators studied by Boldin, Simonova, and Tyurin (1997), are solutions of

$$\hat{\beta}_n(I_p) \in \arg\min_{\beta\in\mathbb{R}^p} s'(Y - X\beta)XX's(Y - X\beta) = \arg\min_{\beta\in\mathbb{R}} SB(\beta), \tag{4.5}$$

and

$$\hat{\beta}_n[(X'X)^{-1}] \in \arg\min_{\beta\in\mathbb{R}^p} s'(Y - X\beta)X(X'X)^{-1}X's(Y - X\beta) = \arg\min_{\beta\in\mathbb{R}} SF(\beta). \tag{4.6}$$

For heteroskedastic independent disturbances, we introduce weighted versions of sign-based estimators that can be more efficient than the basic ones defined in (4.5) or (4.6). Weighted sign-based estimators are sign-based analogues to weighted LAD estimator [see Zhao (2001)]. The weighted LAD estimator is given by

$$\beta_n^{WLAD} = \operatorname*{argmin}_{\beta\in\mathbb{R}^p} \sum_i d_i|y_i - x_i'\beta|. \tag{4.7}$$

The weighted sign-based estimators are solutions of

$$\hat{\beta}_n^{DX} \in \operatorname*{argmin}_{\beta\in\mathbb{R}^p} s'(Y - X\beta)\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'D's(Y - X\beta) \tag{4.8}$$

with

$$\tilde{X} = \begin{Bmatrix} d_1 & 0 & \dots \\ & d_i & \\ 0 & \dots & d_n \end{Bmatrix} X$$

where $(d_i)_{i=1,\dots,n}$ are positive reals. Weighted sign-based estimators that involve optimal estimating functions in the sense of Godambe (2001) are solutions of

$$\hat{\beta}_n^{DX*} \in \operatorname*{argmin}_{\beta\in\mathbb{R}^p} s'(Y - X\beta)X^*(X^{*\prime}X^*)^{-1}X^{*\prime}D's(Y - X\beta) \tag{4.9}$$

where

$$X^* = \begin{Bmatrix} f_1(0|X) & 0 & \dots \\ & f_i(0|X) & \\ 0 & \dots & f_n(0|X) \end{Bmatrix} X$$

with $f_t(0|X), t = 1, \dots, n$, the conditional disturbance density evaluated at zero. The inherent problem of such a class of estimators is to provide good approximations of $f_i(0|X)$'s. Densities of normal distributions can be used.

4.3. Sign-based estimators as GMM estimators

Sign-based estimators have been interpreted in the literature as GMM estimators exploiting the orthogonality condition between the signs and the explanatory variables [see Honore and Hu (2004)]. In our opinion, a strictly GMM interpretation hides the link with the testing theory. That is the reason why we first introduced sign-based estimators as Hodges-Lehmann estimators. The quadratic form (4.4) refers to quite unusual moment conditions. The sign transformation evacuates the unknown parameters that affect the error distribution. It validates nonparametric finite-sample-based inference when mediangale Assumption holds. However, in settings where only the sign-moment condition A1 is satisfied, the GMM interpretation of sign-based estimators still applies and entails useful extensions.

For autocorrelated disturbances, an estimator based on a HAC sign-based statistic $D_S(\beta, \hat{J}_n^{-1})$ can be used:

$$\hat{\beta}_n(\hat{J}_n^{-1}) \in \arg\min_{\beta \in \mathbf{R}^p} s'(Y - X\beta)X[\hat{J}_n(s(Y - X\beta), X)]^{-1}X's(Y - X\beta), \tag{4.10}$$

where $\hat{J}_n^{-1}$ accounts for the dependence among the signs and the explanatory variables. β appears twice, first in the constrained signs, second in the weight matrix. In practice, optimizing (4.10) requires one to invert a new matrix $\hat{J}_n$ for each value of β whereas problem (4.6) only requires one inversion of $X'X$. In practice, this numerical problem may quickly become cumbersome similarly to continuously updating GMM. We advocate to use a two-step method: first, solve (4.6) and obtain $\hat{\beta}_n((X'X)^{-1})$; compute then $\hat{J}_n^{-1}(s(Y - X\hat{\beta}_n((X'X)^{-1}), X)$ and finally solve,

$$\hat{\beta}_n^{2S}(\hat{J}_n^{-1}) \in \arg\min_{\beta \in \mathbf{R}^p} s'(Y - X\beta)X[\hat{J}_n(s(Y - X\hat{\beta}_n), X)]^{-1}X's(Y - X\beta). \tag{4.11}$$

The 2-step estimator is not a Hodges-Lehmann estimator anymore. However, it is still consistent and share some interesting finite-sample properties with classical sign-based estimators. The properties of sign-based estimators are studied in the next section.

5. Some basic properties of sign-based estimators

In this section, both finite and asymptotic properties of sign-based estimators are studied. We demonstrate consistency when the parameter is identified under weaker assumptions than the LAD estimator, which validates the use of sign-based estimators even in settings when the LAD estimator fails to converge. Their finite-sample behavior also presents useful features. They share invariance properties with the LAD estimator and are median-unbiased if the disturbance distribution is symmetric. Finally, sign-based estimators are asymptotically normal.

5.1. Identification and consistency

We show that the sign-based estimators (4.4) and (4.11) are consistent under the following set of assumptions:

Assumption A4 MIXING. *$\{W_t = (y_t, x_t')\}_{t=1,2,\dots}$ is α-mixing of size $-r/(r-1)$ with $r > 1$.*

Assumption A5 BOUNDEDNESS. *$x_t = (x_{1t}, \dots, x_{pt})'$ and $E|x_{ht}|^{r+1} < \Delta < \infty,\ h = 1, \dots, p,\ t = 1, \dots, n,\ \forall n \in \mathbb{N}$.*

Assumption A6 COMPACTNESS. *$\beta \in Int(\Theta)$, where Θ is a compact subset of $\mathbb{R}^p$.*

Assumption A7 REGULARITY OF THE DENSITY.

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

$$P[f_t(0|X) > f_L] > p_1,\ \forall t = 1, \dots, n,\ a.s.$$

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

Assumption A8 POINT IDENTIFICATION CONDITION. *$\forall \delta > 0, \exists \tau > 0$ such that*

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

Assumption A9 UNIFORMLY POSITIVE DEFINITE WEIGHT MATRIX. *$\Omega_n(\beta)$ is symmetric definite positive for all β in Θ.*

Assumption A10 LOCALLY POSITIVE DEFINITE WEIGHT MATRIX NEAR β_0. *$\Omega_n(\beta)$ is symmetric definite positive for all β in a neighborhood of β_0.*

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

Theorem 5.1 CONSISTENCY. *Under model (2.1) with the Assumptions A1, A4-A9, any sign-based estimator of the type,*

$$\hat{\beta}_n(\Omega_n) \in \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}}\, s'(Y - X\beta)X\Omega_n\big(s(y - X\beta), X\big)X's(Y - X\beta), \tag{5.12}$$

or

$$\hat{\beta}_n^{2S}(\Omega_n) \in \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}}\, s'(Y - X\beta)X\hat{\Omega}_n\big(s(y - X\hat{\beta}), X\big)X's(Y - X\beta), \tag{5.13}$$

where $\hat{\beta}$ stands for any (first step) consistent estimator of β, is consistent. $\hat{\beta}_n^{2S}$ defined in equation (5.13) is still consistent if Assumption A9 is replaced by Assumption A10.

Let us interpret precisely Assumptions A4-A10 and compare them to the ones required for LAD and quantile estimator consistency [see Fitzenberger (1997b) and Weiss (1991) for the most general setups]. Assumptions on mixing (A4), compactness (A6) and point identification (A7, A8, A9) are classical. The mixing setup A4 is needed to apply a generic weak law of large numbers [see Andrews (1987) and White (2001)]. It was used by Fitzenberger (1997b) to show LAD and quantile estimator consistency with stationary linearly dependent processes. It covers, among other processes, stationary ARMA disturbances with continuously distributed innovations. Point identification is provided by Assumptions A8 and A7. Assumption A8 is similar to Condition ID in Weiss (1991). Assumption A7 is usual in the LAD estimator asymptotics.[5] It is analogous to Fitzenberger (1997b)'s conditions (ii.b and c) and Weiss (1991)'s CD condition. It implies that there is enough variation around zero to identify the median. It restricts the setup for some 'bounded' heteroskedasticity in the disturbance process but not in the usual (variance-based) way. Indeed, so-called

[5] Assumption A7 can be slightly relaxed covering error terms with mass point if the objective function involves randomized signs instead of usual signs

diffusivity, $\frac{1}{2f(0)}$, can be seen as an alternative measure of dispersion adapted to median-unbiased estimators. It measures the vertical spread of a density rather than its horizontal spread and is involved in Cramér-Rao-type lower bound for median-unbiased estimators [see Sung, Stangenhaus, and David (1990) and So (1994)]. Besides, in Assumptions A9 and A10, the weight matrix Ω_n is supposed to be invertible for estimators obtained in one step whereas only a local invertibility is needed for two-step sign-estimators. One difference with the LAD asymptotic properties relies on Assumption A5. For sign consistency, only the second-order moments of x_t have to be finite, which differs from Fitzenberger (1997b) who supposed the existence of at least third-order moments. And above all, we do not assume the existence of second-order moments on the disturbances u_t. Indeed, the disturbances appear in the objective function only through their sign transforms which possess finite moments up to any order. Consequently, no additional restriction should be imposed on the disturbance process (in addition to regularity conditions on the density). Those points will entail a more general CLT than the one stated for the LAD/quantile estimators in Fitzenberger (1997b) and Weiss (1991).

5.2. Unbiasedness and equivariance

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

Proposition 5.2 EQUIVARIANCE. *If $\hat{\beta}(y, X, u)$ is a solution of (4.4), then*

$$\hat{\beta}(\lambda y, X, u) = \lambda \hat{\beta}(y, X, u), \qquad \forall \lambda \in \mathbb{R} \tag{5.14}$$

$$\hat{\beta}(y + X\gamma, X, u) = \hat{\beta}(y, X, u) + \gamma, \qquad \forall \gamma \in \mathbb{R}^p \tag{5.15}$$

$$\hat{\beta}(y, XA, u) = A^{-1}\hat{\beta}(y, X, u), \quad \textit{for any nonsingular } k \times k \textit{ matrix } A. \tag{5.16}$$

To prove this property, it is sufficient to write down the different optimization problems. Equation (5.14) states that $\hat{\beta}$ is scale invariant: if y is rescaled by a certain factor, $\hat{\beta}$ is rescaled by the same one. Equation (5.15) states that $\hat{\beta}$ is location invariant, while (5.16) states a reparameterization invariance with respect to the design matrix: the transformation on $\hat{\beta}$ is given by the inverse of the reparameterization scheme.

Moreover, if the disturbance distribution is assumed to be symmetric then sign-estimators are median unbiased.

Proposition 5.3 MEDIAN UNBIASEDNESS. *If $u \sim -u$, then any sign-based estimator $\hat{\beta}$ solution of minimization problem (4.4) is median unbiased, that is,*

$$Med(\hat{\beta} - \beta_0) = 0$$

where β_0 is the true value.

5.3. Asymptotic normality

Sign-based estimators are asymptotically normal. This also holds under weaker assumptions than the ones needed for LAD estimator asymptotic normality. Sign-based estimators are specially adapted for heavy-tailed disturbances that may not possess finite variance. The assumptions we need are the following ones.

Assumption A11 UNIFORMLY BOUNDED DENSITIES. $\exists f_U < +\infty$ *such that* , $\forall n \in \mathbb{N}, \forall \lambda \in \mathbb{R}$,

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

Under the conditions A1, A4, A5 and A11, we can define $L(\beta)$, the derivative of the limiting objective function at β:

$$L(\beta) = \lim_{n\to\infty} \frac{1}{n} \sum_t E\big[x_t x_t' f_t\big(x_t'(\beta - \beta_0) | x_1, \ldots, x_n\big)\big] = \lim_{n\to\infty} L_n(\beta). \tag{5.17}$$

where

$$L_n(\beta) = \frac{1}{n} \sum_t E\big[x_t x_t' f_t\big(x_t'(\beta - \beta_0) | x_1, \ldots, x_n\big)\big]. \tag{5.18}$$

The other assumptions are merely used to show asymptotic normality.

Assumption A12 MIXING WITH $r > 2$. $\{W_t = (y_t, x_t')\}_{t=1,2,\ldots}$ *is α-mixing of size $-r/(r-2)$ with $r > 2$.*

Assumption A13 DEFINITE POSITIVENESS OF L_n. *$L_n(\beta_0)$ is positive definite uniformly in n.*

Assumption A14 DEFINITE POSITIVENESS OF J_n. *The matrix $J_n = E[\frac{1}{n}\sum_{t,s}^{n} s(u_t)x_t x_s' s(u_s)]$ is positive definite uniformly in n and converges to a definite positive symmetric matrix J.*

Then, we have the following result.

Theorem 5.4 ASYMPTOTIC NORMALITY. *Under the conditions for consistency (A1, A4-A9), and A12-A14, we have:*

$$S_n^{-1/2}\sqrt{n}(\hat{\beta}_n(\Omega_n) - \beta_0) \xrightarrow{d} N(0, I_p) \tag{5.19}$$

where

$$S_n = [L_n(\beta_0)\Omega_n L_n(\beta_0)]^{-1} L_n(\beta_0)\Omega_n J_n \Omega_n L_n(\beta_0)[L_n(\beta_0)\Omega_n L_n(\beta_0)]^{-1}$$

and

$$L_n(\beta_0) = \frac{1}{n}\sum_t E\big[x_t x_t' f_t(0|x_1, \ldots, x_n)\big]. \tag{5.20}$$

Remark that when $\Omega_n = \hat{J}_n^{-1}$, we have

$$[L_n(\beta_0)\hat{J}_n^{-1}L_n(\beta_0)]^{-1/2}\sqrt{n}(\hat{\beta}_n(\hat{J}_n^{-1}) - \beta_0) \xrightarrow{d} \mathrm{N}(0, I_p). \tag{5.21}$$

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

Note again that we do not require that the disturbance process variance be finite. We only assume that the second-order moments of X are finite and the mixing property of $\{W_t,\ t = 1, \ldots\}$ holds. This differs from usual assumptions for LAD asymptotic normality.[6] This difference comes from the fact that absolute values of the disturbance process

[6] See Fitzenberger (1997b) for the derivation of the LAD asymptotics in a similar setup and Koenker-Bassett(1978) or Weiss (1991) for a derivation of the LAD asymptotics under sign independence

are replaced in the objective function by their signs. Since signs possess finite moments at any order, one sees easily that a CLT can be applied without any further restriction. Consequently, asymptotic normality, such as consistency, holds for heavy-tailed disturbances that may not possess finite variance. This is an important theoretical advantage of sign-based rather than absolute value-based estimators and, *a fortiori*, rather than least squares estimators. Estimators for which asymptotic normality holds on bounded asymptotic variance assumption (for example OLS) are not accurate in heavy-tail settings because the variance is not a measure of dispersion adapted to those settings. Estimators, for which the asymptotic behavior relies on other measures of dispersion, like the diffusivity, help one out of trouble.

The form of the asymptotic covariance matrix simplifies under stronger assumptions. When the signs are mutually independent conditional on X [mediangale Assumption A2], both $\hat{\beta}_n((X'X)^{-1})$ and $\beta(\hat{J}_n^{-1})$ are asymptotically normal with variance

$$S_n = \quad [L_n(\beta_0)]^{-1} E\left[(1/n) \sum_{t=1}^{n} x_t x_t'\right] [L_n(\beta_0)]^{-1}.$$

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

$$S_n = \frac{1}{4f(0)^2} E(x_t x_t')^{-1}. \tag{5.22}$$

In the general case, $f_t(0)$ is a nuisance parameter even if condition A11 implies that it can be bounded.

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

5.4. Asymptotic or projection-based sign-confidence intervals?

In section 4, we introduced sign-based estimators as Hodges-Lehmann estimators associated with sign-based statistics. By linking them with GMM settings, we then derived asymptotic normality. We stressed that sign-based estimator asymptotic normality holds under weaker assumptions than the ones needed for the LAD estimator. Therefore, sign-based estimator asymptotic normality enables one to construct asymptotic tests and confidence intervals. Thus, we have two ways of making inference with signs: we can use the Monte Carlo (finite-sample) based method described in Coudin and Dufour (2005a)- see subsection 2.2- and the classical asymptotic method. Let us list here the main differences between them. Monte Carlo inference relies on the pivotality of the sign-based statistic. The derived tests are valid (with controlled level) for any sample size if the mediangale Assumption A2 holds. When only the sign moment condition A1 holds, the Monte Carlo inference remains asymptotically valid. Asymptotic test levels are controlled. Besides, in simulations, the Monte Carlo inference method appears to perform better in small samples than classical asymptotic methods, even if its use is only asymptotically justified [see Coudin and Dufour (2005a)]. Nevertheless, that method has an important drawback: its computational complexity. On the contrary, classical asymptotic methods which yield tests with controlled asymptotic level under the sign moment condition A1 may be less time consuming. The choice between both is mainly a question of computational capacity. We point out that classical asymptotic inference greatly relies on the way the asymptotic covariance matrix, that depends on unknown parameters (densities at zero), is treated. If the asymptotic covariance matrix is estimated thanks to a simulation-based method (such as the bootstrap) then the time argument does not hold anymore. Both methods would be of the same order of computational complexity.

6. Simulation study

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

6.1. Setup

We use estimators derived from the sign-based statistics $D_S(\beta, (X'X)^{-1})$ and $D_S(\beta, \hat{J}_n^{-1})$ when a correction is needed for linear serial dependence. We consider a set of general

DGP's to illustrate different classical problems one may encounter in practice. We use the following linear regression model:

$$y_t = x_t'\beta_0 + u_t, \tag{6.1}$$

where $x_t = (1, x_{2,t}, x_{3,t})'$ and β_0 are 3×1 vectors. We denote the sample size n. Monte Carlo studies are based on M generated random samples. Table 12 presents the cases considered.

In a first group of examples (A1-A4), we consider classical independent cases with bounded heterogeneity. In a second one (B5-B8), we look at processes involving large heteroskedasticity so that some of the estimators we consider may not be asymptotically normal neither consistent anymore. Finally, the third group (C9-C11) is dedicated to autocorrelated disturbances. We wonder whether the two-step $SHAC$ sign-based estimator performs better in small samples than the non-corrected one.

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

Table 12. Simulated models.

CASE A1:	Normal HOM:	$(x_{2,t}, x_{3,t}, u_t)' \overset{i.i.d}{\sim} \mathcal{N}(0, I_3),\ t = 1, \ldots, n$
CASE A2:	Normal HET:	$(x_{2,t}, x_{3,t}, \tilde{u}_t)' \overset{i.i.d}{\sim} \mathcal{N}(0, I_3)$ $u_t = min\{3, max[0.21, \lvert x_{2,t}\rvert]\} \times \tilde{u}_t,\ t = 1, \ldots, n$
CASE A3:	Dep.-HET, $\rho_x = .5$:	$x_{j,t} = \rho_x x_{j,t-1} + \nu_t^j,\ j = 1, 2,$ $u_t = min\{3, max[0.21, \lvert x_{2,t}\rvert]\} \times \nu_t^u,$ $(\nu_t^2, \nu_t^3, \nu_t^u)' \overset{i.i.d}{\sim} \mathcal{N}(0, I_3),\ t = 2, \ldots, n$ ν_1^2 and ν_1^3 chosen to insure stationarity.
CASE A4:	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, $t = 1, \ldots, n.$
CASE B5:	Cauchy dist.:	$(x_{2,t}, x_{3,t})' \sim \mathcal{N}(0, I_2),$ $u_t \overset{i.i.d.}{\sim} \mathcal{C}, x_t, u_t,$ independent, $t = 1, \ldots, n.$
CASE B6:	Stoc. Volat.:	$(x_{2,t}, x_{3,t})' \overset{i.i.d.}{\sim} \mathcal{N}(0, I_2), u_t = exp(w_t/2)\epsilon_t$ with $w_t = 0.5 w_{t-1} + v_t$, 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, $t = 1, \ldots, n.$
CASE B7:	Nonstat. GARCH(1,1):	$(x_{2,t}, x_{3,t}, \epsilon_t)' \overset{i.i.d.}{\sim} \mathcal{N}(0, I_3),\ t = 1, \ldots, n,$ $u_t = \sigma_t \epsilon_t,\ \sigma_t^2 = 0.8 u_{t-1}^2 + 0.8 \sigma_{t-1}^2.$
CASE B8:	Exp. Var.:	$(x_{2,t}, x_{3,t}, \epsilon_t)' \overset{i.i.d.}{\sim} \mathcal{N}(0, I_3), u_t = exp(.2t)\epsilon_t.$
CASE C9:	AR(1)-HOM, $\rho_u = .5$:	$(x_{2,t}, x_{3,t}, \nu_t^u)' \sim \mathcal{N}(0, I_3), t = 2, \ldots, n,$ $u_t = \rho_u u_{t-1} + \nu_t^u,$ $(x_{2,1}, x_{3,1})' \sim \mathcal{N}(0, I_2), \nu_1^u$ insures stationarity.
CASE C10:	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.
CASE C11:	AR(1)-HOM, $\rho_u = .9$:	$(x_{2,t}, x_{3,t}, \nu_t^u)' \sim \mathcal{N}(0, I_3), t = 2, \ldots, n,$ $u_t = \rho_u u_{t-1} + \nu_t^u,$ $(x_{2,1}, x_{3,1})' \sim \mathcal{N}(0, I_2), \nu_1^u$ insures stationarity.

6.2. Bias and RMSE

We give biases and RMSE of each parameter of interest in Table 13 and we report a norm of these three values. $n = 50$ and $S = 1000$. These results are unconditional on X.

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

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

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

To conclude, sign-based estimators are robust estimators much less sensitive than the LAD estimator to various debalanced schemes in the explanatory variables and to heteroskedasticity. They are particularly adequate when an amount an heteroskedasticity or nonlinear dependence is suspected in the error term, even if the error term fails to be stationary. Finally, the HAC correction does not seem to increase the performance of the estimator. Nevertheless, it does for tests. We show in Coudin and Dufour (2005a) that using a HAC-corrected statistic allows for the asymptotic validity of the Monte Carlo inference method and improves the test performance in small samples.

Table 13. Simulated bias and RMSE.

$n = 50, \quad S = 1000$		OLS Bias	OLS RMSE	LAD Bias	LAD RMSE	SF Bias	SF RMSE	2SSHAC Bias	2SSHAC RMSE
Case A1:	β_0	.003	.142	.002	.179	.002	.179	.004	.178
$\rho_\epsilon = \rho_x = 0$,	β_1	.003	.149	.006	.184	.004	.182	.004	.182
HOM	β_2	-.002	.149	-.007	.186	-.006	.185	-.007	.183
	$\|\|\beta\|\|$*	**.004**	**.254**	**.009**	**.316**	**.007**	**.315**	**.009**	**.313**
Case A2:	β_0	-.003	.136	.000	.090	-.000	.089	-.000	.089
$\rho_\epsilon = \rho_x = 0$,	β_1	-.0135	.230	-.006	.218	-.010	.218	-.010	.218
HET	β_2	.002	.142	-.001	.095	-.001	.092	-.001	.092
	$\|\|\beta\|\|$	**.014**	**.303**	**.007**	**.254**	**.010**	**.253**	**.010**	**.253**
Case A3:	β_0	.022	.167	.018	.108	.025	.107	.023	.107
$\rho_\epsilon = 0,\ \rho_x = .5$,	β_1	-1.00	.228	.005	.215	.003	.214	.002	.215
HET	β_2	.001	.150	.005	.105	.007	.104	.007	.105
	$\|\|\beta\|\|$	**.022**	**.320**	**.019**	**.263**	**.026**	**.261**	**.024**	**.262**
Case A4:	β_0	-.001	.174	.007	.2102	.010	.2181	.008	.2171
$x_2 \sim \mathcal{B}(1, .3)$,	β_1	-.016	.313	-.011	.375	-.021	.396	-.021	.394
$x_3 \sim \mathcal{N}(0, .01^2)$	β_2	-.100	14.6	.077	18.4	.014	7.41	.049	7.40
	$\|\|\beta\|\|$	**.101**	**14.6**	**.078**	**18.5**	**.027**	**7.42**	**.054**	**7.41**
Case B5:	β_0	16.0	505	.001	.251	.004	.248	.003	.248
$x_2, x_3 \sim \mathcal{N}(0, I_2)$,	β_1	-3.31	119	.015	.264	.020	.265	.020	.265
$u \sim \mathcal{C}$	β_2	-2.191	630	.000	.256	.003	.258	.001	.258
	$\|\|\beta\|\|$	**26.0**	**817**	**.015**	**.445**	**.021**	**.445**	**.020**	**.445**
Case B6:	β_0	-.908	29.6	-1.02	27.4	.071	2.28	.083	2.28
Stoch. Volat.	β_1	2.00	37.6	3.21	68.4	.058	2.38	.069	2.39
	β_2	1.64	59.3	2.59	91.8	-.101	2.30	-.089	2.29
	$\|\|\beta\|\|$	**2.73**	**76.2**	**4.25**	**118**	**.136**	**4.02**	**.139**	**4.02**
Case B7:	β_0	-127	3289	-.010	7.85	-.008	3.16	-.028	3.17
GARCH(1,1)	β_1	-81.4	237	.130	11.2	-.086	3.80	-.086	3.823
	β_2	-31.0	1484	-.314	12.0	-.021	3.606	-.009	3.630
	$\|\|\beta\|\|$	**154**	**4312**	**.340**	**18.2**	**.089**	**6.12**	**.091**	**6.15**
Case B8:	β_0	$< -10^{10}$	$> 10^{10}$	$< -10^{9}$	$> 10^{10}$	.312	5.67	.307	5.67
$u_t = exp(t)\epsilon_t$	β_1	$> 10^{10}$	$> 10^{10}$	$> 10^{9}$	$> 10^{10}$	.782	5.40	.863	5.46
	β_2	$< -10^{10}$	$> 10^{10}$	$< -10^{9}$	$> 10^{10}$	.696	5.52	.696	5.55
	$\|\|\beta\|\|$	$\mathbf{> 10^{10}}$	$\mathbf{> 10^{10}}$	$\mathbf{> 10^{10}}$	$\mathbf{> 10^{10}}$	**1.09**	**9.58**	**1.15**	**9.63**
Case C9:	β_0	.005	.279	.001	.308	.003	.309	.004	.311
$\rho_\epsilon = .5, \rho_x = 0$,	β_1	-.002	.163	-.005	.201	-.004	.200	-.005	.199
HOM	β_2	.001	.165	-.004	.204	.003	.198	.002	.198
	$\|\|\beta\|\|$	**.006**	**.363**	**.007**	**.420**	**.006**	**.418**	**.006**	**.419**
Case C10:	β_0	-.013	.284	-.010	.315	-.015	.314	-.014	.314
$\rho_\epsilon = .5, \rho_x = .5$,	β_1	-.009	.182	-.009	.220	-.011	.218	-.011	.219
HET	β_2	.008	.189	.011	.222	.007	.215	.007	.215
	$\|\|\beta\|\|$	**.018**	**.387**	**.018**	**.444**	**.020**	**.439**	**.019**	**.439**
Case C11:	β_0	.070	1.23	-.026	.308	.058	1.26	.053	1.27
$\rho_\epsilon = .9, \rho_x = 0$,	β_1	-.000	.268	.005	.214	-.005	.351	-.008	.354
HOM	β_2	.001	.273	-.004	.210	.002	.361	-.001	.361
	$\|\|\beta\|\|$	**.070**	**1.29**	**.027**	**.430**	**.059**	**1.36**	**.054**	**1.37**

* ||.|| stands for the euclidian norm.

7. Illustrations

In this section, we go back to the two illustrations presented in Coudin and Dufour (2005a) where sign-based tests were derived, with now estimation in mind. The first application is dedicated to estimate a drift on the Standard and Poor's Composite Price Index (S&P), 1928-1987. In the second one, we search a robust estimate of the rate of β convergence between output levels across U.S. States during the 1880-1988 period using Barro and Sala-i Martin (1991) data.

7.1. Drift estimation with stochastic volatility in the error term

We estimate a constant and 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 Valéry and Dufour (2004) to fit a stochastic volatility model. Here, we are interested in robust **estimation** without modeling the volatility in the disturbance process. The data set consists in a series of 16,127 daily observations of SP_t, then converted in price movements, $y_t = 100[\log(SP_t) - \log(SP_{t-1})]$ and adjusted for systematic calendar effects. We consider a model involving a constant and a drift,

$$y_t = a + bt + u_t, \quad t = 1, \dots, 16127, \tag{7.2}$$

and we allow that $\{u_t\}_{t=1,\dots,16127}$ exhibits stochastic volatility or nonlinear heteroskedasticity of unknown form. White and Breush-Pagan tests for heteroskedasticity both reject homoskedasticity at 1%.[7]

We compute both the basic SF sign-based estimator and the $SHAC$ version with the two-step method. They are compared with the LAD and OLS estimates. Then, we redo a similar experiment on two subperiods: on the year 1929 (291 observations) and the last 90 days of 1929, which roughly corresponds to the four last months of 1929 (90 observations). Due to the financial crisis, one may expect data to involve an extreme amount of heteroskedasticity in that period of time. We wonder at which point that heteroskedasticy

[7] See Coudin and Dufour (2005a): White: 499 (p-value=.000) ; BP: 2781 (p-value=.000).

can bias the subsample estimates. The Wall Street krach occurred between October, 24th (*Black Thursday*) and October, 29th (*Black Tuesday*). Hence, the second subsample corresponds to the period just before the krach (September), the krach period (October) and the early beginning of the Great Depression (November and December). Heteroskedasticity tests reject homoskedasticity for both subsamples.[8]

In Table 14, we report estimates and recall the 95% confidence intervals for a and b obtained by the finite-sample sign-based method (SF and $SHAC$);[9] and by moving block bootstrap (LAD and OLS). The entire set of sign-based estimators is reported, *i.e.*, all the minimizers of the sign objective function.

Table 14. Constant and drift estimates.

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

* Interval of admissible estimators (minimizers of the sign objective function).
** 95% confidence intervals.

First, we note that the OLS estimates are importantly biased and are greatly unreliable

[8] 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.
[9] see Coudin and Dufour (2005a)

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

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

7.2. A robust sign-based estimate of β convergence across US States.

One field suffering from both a small number of observations and possibly very heterogeneous data is cross-sectional regional data sets. Least squares methods may be misleading because a few outlying observations may drastically influence the estimates. Robust methods are greatly needed in such cases. Sign-based estimators are robust (in a statistical sense) and are naturally associated with a finite-sample inference. In the following, we examine sign-based estimates of the rate of β convergence between output levels across U.S. States between 1880 and 1988 using Barro and Sala-i Martin (1991) data.

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

$$(1/T)\ln(y_{i,t}/y_{i,t-T}) \;=\; a - [\ln(y_{i,t-T})] \times [(1 - e^{-\beta T})/T] + x_i'\delta + \epsilon_i^{t,T}, \qquad (7.3)$$

$$i = 1, \ldots, 48,\ T = 8, 10 \text{ or } 20,$$
$$t = 1900, 1920, 1930, 1940, 1950, 1960, 1970, 1980, 1988.$$

Their *basic equation* does not include other explanatory variables but they also consider a specification with regional dummies. The *basic equation* model assumes that 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.

Their results suggest a β convergence at a rate somewhat above 2% a year but their estimates are not stable across subperiods, and vary greatly from -.0149 to .0431 (for the *basic equation*). This instability is expected because of the succession of troubles and growth periods in the last century. However, they may also be due to particular observations behaving like outliers and influencing the least squares estimates.[10] These two effects are probably combined. We wonder which part of that variability is really due to business cycles and which part is only due to the nonrobustness of least squares methods. Further, we would like to have a stable estimate of the rate of convergence at steady state. For this, we use robust sign-based estimation with $D_S(\beta, (X'X)^{-1})$. We consider the following linear equation:

$$(1/T)\ln(y_{i,t}/y_{i,t-T}) = a + \gamma[\ln(y_{i,t-T})] + x_i'\delta + \epsilon_i^{t,T}, \tag{7.4}$$
$$i = 1, \ldots, 48,\ T = 8, 10 \text{ or } 20,$$
$$t = 1900, 1920, 1930, 1940, 1950, 1960, 1970, 1980, 1988.$$

where x_i account for regional dummies when included, and we compute Hodges-Lehmann estimate for $\beta = -(1/T)\ln(\gamma T + 1)$ for both specifications. We also provide 95%-level projection-based CI, asymptotic CI and projection-based p-value functions for the parameter of interest. Results are presented in Table 15 where Barro and Sala-i Martin (1991)

[10] A survey of potential data problem is performed and regression diagnostics are summarized in Table 16 in the Appendix B.1. It suggests the presence of highly influential observations in all the periods but one. Outliers are clearly identified in periods 1900-1920, 1940-1950, 1950-1960, 1970-1980 and 1980-1988.

NLLS results are reported.

Sign estimates are more stable than least squares ones. They vary between $[-.0147, .0364]$ whereas least squares estimates vary between $[-.0149, .0431]$. This suggests that at least 12% of the least squares estimates variability between sub-periods are only due to the nonrobustness of least squares methods. In all cases but two, sign-based estimates are lower (in absolute values) than the NLLS ones. Consequently, we incline to a lower value of the stable rate of convergence.

In graphics 28(a)-30(f) [see Appendix B.2], projection-based p-value functions and optimal concentrated sign-statistics are presented for each *basic equation* over the period 1880-1988. The optimal concentrated sign-based statistic reports the minimal value of D_S for a given β (letting a varying). The projection-based p-value function is the maximal simulated p-value for a given β over admissible values of a. Those functions enable us to perform tests on β. 95% projection based confidence intervals for β presented in Table 15 are obtained by cutting the p-value function with the $p = .05$ line. The sign estimate reaches the highest p-value. Remark that contrary to asymptotic methods, the estimator is not at the middle point of any confidence interval. Besides, the p-value function gives some hint on the degree of precision. The β parameter seems precisely estimated in the period 30-40 [see graphic 29(b)], whereas in the period 80-88, the same parameter is less precisely estimated and the p-value function leads to a wider confidence intervals [see graphic 30(f)].

8. Conclusion

In this paper, we introduce inference tools that can be associated with the Monte Carlo based system presented in Coudin and Dufour (2005a): the p-value function (and its individual projected versions) which gives a visual summary of all the inference available on a particular parameter, and Hodges-Lehmann-type sign-based estimators. The p-value function associates to each value of the parameter the degree of confidence one may have in that particular value. It extends the confidence distribution concept to multidimensional parameters and relies on a reinterpretation of the Fisher fiducial distributions. The para-

Table 15. Regressions for personal income across U.S. States, 1880-1988.

Period	*Basic equation*		*Equation with regional dummies*	
	β^{SIGN}	β^{NLLS} * **	β^{SIGN}	β^{NLLS} * **
1880-1900	.0012 *[-.0068, .0123]**	.0101 *[.0058, .0532]***	.0016 *[-.0123, .0211]*	.0224 *[.0146, .0302]*
1900-1920	.0184 *[.0092, .0313]*	.0218 *[.0155, .0281]*	.0163 *[-.0088, .1063]*	.0209 *[.0086, .0332]*
1920-1930	-.0147 *[-.0301, .0018]*	-.0149 *[-.0249, -.0049]*	-.0002 *[-.0463, .0389]*	-.0122 *[-.0267, .0023]*
1930-1940	.0130 *[.0043, .0234]*	.0141 *[.0082, .0200]*	.0152 *[-.0189, .0582]*	.0127 *[.0027, .0227]*
1940-1950	.0364 *[.0291, .0602]*	.0431 *[.0372, .0490]*	.0174 *[.0083, .0620]*	.0373 *[.0314, .0432]*
1950-1960	.0195 *[.0084, .0352]*	.0190 *[.0121, .0259]*	.0140 *[-.0044, .0510]*	.0202 *[.0100, .0304]*
1960-1970	.0289 *[.0099, .0377]*	.0246 *[.0170, .0322]*	.0230 *[-.0112, .0431]*	.0131 *[.0047, .0215]*
1970-1980	.0181 *[.0021, .0346]*	.0198 *[-.0315, .0195]*	.0172 *[-.0131, .0739]*	.0119 *[-.0273, .0173]*
1980-1988	-.0081 *[-.0552, .0503]*	-.0060 *(.0130)*	-.0059 *[-.0472, .1344]*	-.0050 *(.0114)*

* Projection-based 95% CI.
** Asymptotic 95% CI.
*** Columns 2 and 4 are taken from Barro and Sala-i Martin (1991).

meter values the less rejected by tests (given the sample realization and the sample size) constitute Hodges-Lehmann sign-based estimators. Those estimators are associated with the highest p-value. Hence, they are derived without referring to asymptotic conditions through the analogy principle. However, they turn out to be equivalent (in probability) to usual GMM estimators based on signs. We then present general properties of sign-based estimators (invariance, median unbiasedness) and the conditions under which consistency and asymptotic normality hold. In particular, we show that sign-based estimators do require less assumptions on moment existence of the disturbances than usual LAD asymptotic theory. Simulation studies indicate that the proposed estimators are accurate in classical setups and more reliable than usual methods (LS, LAD) when arbitrary heterogeneity or nonlinear dependence is present in the error term even in cases that may cause LAD or OLS consistency failure. Despite the programming complexity of sign-based methods, we recommend combining sign-based estimators to the Monte Carlo sign-based method of inference when an amount of heteroskedasticity is suspected in the data and when the number of available observations is small. We present two illustrative applications of such cases. In the first one, we estimate a drift parameter on the Standard and Poor's Composite Price Index, using the 1928-1987 period and various shorter subperiods. In the second one, we provide robust estimates for the β convergence between the levels of per capita personal income across U.S. States occurred between 1880 and 1988.

Appendix

A. Proofs

A.1. Proof of Proposition 4.1

We show that the sets $M1$ and $M2$ are equal with probability one. First, we show that if $\hat{\beta} \in M2$ then it belongs to $M1$. Second, we show that if $\hat{\beta}$ does not belong to $M2$, neither it belongs to $M1$.
If $\hat{\beta} \in M2$ then,

$$D_S(\hat{\beta}) \leq D_S(\beta),\ \forall \beta \in \mathbb{R}^p, \tag{A.1}$$

hence

$$\bar{F}\big(D_S(\hat{\beta})\big) \leq \bar{F}\big(D_S(\beta)\big) = 1, \quad \forall \beta \in \mathbb{R}^p \tag{A.2}$$

and $\hat{\beta}$ maximizes the p-value. Conversely, if $\hat{\beta}$ does not belong to $M1$, there is a non negligible Borel set, say A, such that $D_S(\beta) < D_S(\hat{\beta})$ on A for some β. Then, as $\bar{F}(x)$ is an increasing function and A is non negligible,

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

hence,

$$1 - \bar{F}\big(D_S(\beta)\big) \ > \ 1 - \bar{F}\big(D_S(\hat{\beta})\big).$$

The latter expression can be written in terms of p-values:

$$\mathrm{p}(\beta) > \mathrm{p}(\hat{\beta}), \tag{A.4}$$

and $\hat{\beta}$ does not belong to $M2$.

A.2. Proof of Theorem 5.1

We consider the stochastic process $W = \{W_t = (y_t,\ x_t') : \Omega \rightarrow \mathbb{R}^{p+1}\}_{t=1,2,...}$ defined on the probability space $(\Omega, \mathcal{F}, P)$. We denote

$$q_t(W_t, \beta) \quad = \quad \big(q_{t1}(W_t, \beta), \ldots, q_{tp}(W_t, \beta)\big)'$$

$$= \left(s(y_t - x_t'\beta)x_{t1}, \ldots, s(y_t - x_t'\beta)x_{tp}\right)', \ t = 1, \ldots, n.$$

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

Pointwise convergence. The mixing property A4 on W is exported to $\{q_{tk}(W_t,\beta),\ k = 1,\ldots,p\}_{t=1,2,\ldots}$. Hence, $\forall\beta \in \Theta$, $\forall k = 1,\ldots,p$, $\{q_{tk}(W_t,\beta)\}$ is an $\alpha-$mixing process of size $r/(1-r)$. Moreover, condition A5 entails $E|q_{tk}(W_t,\beta)|^{r+\delta} < \infty$ for some $\delta > 0$, for all $t \in \mathbb{N}$, $k = 1,\ldots,p$. Hence, we can apply Corollary 3.48 of White (2001) to $\{q_{tk}(W_t,\beta)\}_{t=1,2,\ldots}$. It follows $\forall\beta \in \Theta$,

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

Uniform Convergence. We check conditions A1, A6, B1, B2 of Andrews (1987)'s generic weak law of large numbers (GWLLN). A1 and B1 are our conditions A6 and A4. Then, Andrews defines

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

where $B(\beta,\rho)$ is the open ball around β of radius ρ. His condition B2 requires that $q_{tk}^{H}(W_t,\beta,\rho)$, $q_{Ltk}(W_t,\beta,\rho)$ and $q_{tk}(W_t)$ are random variables; $q_{tk}^{H}(.,\beta,\rho)$, $q_{Ltk}(.,\beta,\rho)$ are measurable functions from $(\Omega,\mathcal{P},\mathcal{F})$ to $(\mathbb{R},\mathcal{B})$, $\forall t,\ \beta \in \Theta,\ \rho$, where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}$ and finally, that $\sup_t Eq_{tk}(W_t)^{\xi} < \infty$ with $\xi > r$. Those points are derived from the mixing condition A4 and condition A5 which insures measurability and provides bounded arguments.

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

w.r.t. μ, for each $\beta^* \in \Theta$, $q_{tk}(W_t,\ \beta)$ is measurable w.r.t. the Borel measure for each t and each $\beta \in \Theta$, and $\int \sup_{t\geq 0,\ \beta\in\Theta} |q_{tk}(W,\ \beta)| p_t(w) d\mu(w) < \infty$. As u_t is continuously distributed uniformly in t [Assumption A7 (2)], we have $P_t[u_t = x_t\beta] = 0,\ \forall\beta$, uniformly in t. Then, q_{tk} is continuous in β everywhere except on a P_t-negligeable set. Finally, since q_{tk} is L_1-bounded and uniformly integrable, condition A6 holds.

The generic law of large numbers (GWLLN) implies:

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

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

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

$$\lim_{n\to\infty} \frac{1}{n}\sum_{t} E[s(u_t - x_t'\beta)x_{tk}] \text{ is continuous on } \Theta, k = 1, \ldots, p. \tag{A.5}$$

Let us define

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

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

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

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

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

$$P\left[\sum_{k}[Q_n^k(\beta_n)]^2 < \sum_{k}[Q_n^k(0)]^2 + \epsilon/3\right] \geq 1 - \delta. \tag{A.6}$$

Uniform weak convergence of Q_n^k to Q_n^{Ek} at β_n implies:

$$[Q_n^{Ek}(\beta_n)]^2 < [Q_n^k(\beta_n)]^2 + \epsilon/3p, \; k = 1, \ldots, p, \text{ with probability approaching one as } n \to \infty, \tag{A.7}$$

hence,

$$\sum_k [Q_n^{Ek}(\beta_n)]^2 < \sum_k [Q_n^k(\beta_n)]^2 + \epsilon/3, \text{ with probability approaching one as } n \to \infty. \tag{A.8}$$

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

$$\sum_k [Q_n^k(0)]^2 < \sum_k [Q_n^{Ek}(0)]^2 + \epsilon/3, \text{ with probability approaching one as } n \to \infty. \tag{A.9}$$

Using (A.8), (A.6) and (A.9) in turn, this entails

$$\sum_k [Q_n^{Ek}(\beta_n)]^2 < \sum_k [Q_n^{Ek}(0)]^2 + \epsilon, \text{ with probability approaching one as } n \to \infty. \tag{A.10}$$

This holds for any ϵ, with probability approaching one. Let N be any open subset of Θ containing 0. As $\Theta \cap \mathrm{N}^c$ is compact and $\lim_n \sum_k [Q_n^{*k}(\beta)]^2$ is continuous (A.5),

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

Provided that 0 is the unique minimizer, we have:

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

Hence, setting

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

it follows that, with probability close to one,

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

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

For **identification**, the uniqueness of the minimizer of the sign-objective function is insured by the set of identification conditions A1, A8, A7, A9. These conditions and consequently the proof, are close to those of Weiss (1991) and Fitzenberger (1997b) for the LAD and quantile estimators. We wish to show that the limit problem does not admit another solution. When $\Omega_n(\beta)$ defines a norm for each β (condition A9), this assertion is equivalent to

$$\lim_{n\to\infty} E\left[\frac{1}{n}\sum_t s(u_t - x_t'\delta)x_i\right] = 0 \Rightarrow \delta = 0,\ \delta \in \mathbb{R}^p, \tag{A.11}$$

and

$$\lim_{n\to\infty}\left|E\left[\frac{1}{n}\sum_t s(u_t - x_t'\delta)x_t'\delta\right]\right| = 0 \Rightarrow \delta = 0,\ \delta \in \mathbb{R}^p. \tag{A.12}$$

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

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

Note that

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

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

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

Then,

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

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

Remark that each term in this sum is negative. Hence, $s(E[A(\delta)]) \le 0$ and $|E[A(\delta)]| = -E[A(\delta)]$, and

$$\begin{aligned}
|E(A)| &= E\left[\frac{2}{n}\sum x_t'\delta I_{\{|x_t'\delta|>\tau\}}\left(I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta} f_t(u|x_1,\ldots,x_n)du\right.\right.\\
&\quad \left.\left.-I_{\{x_t'\delta\le 0\}}\int_{x_t'\delta}^0 f_t(u|x_1,\ldots,x_n)du\right)\right]\\
&\quad +E\left[\frac{2}{n}\sum x_t'\delta I_{\{|x_t'\delta|\le\tau\}}\left(I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta} f_t(u|x_1,\ldots,x_n)du\right.\right.\\
&\quad \left.\left.-I_{\{x_t'\delta\le 0\}}\int_{x_t'\delta}^0 f_t(u|x_1,\ldots,x_n)du\right)\right]\\
&\ge E\left[\frac{2}{n}\sum I_{\{|x_t'\delta|>\tau\}}\left(x_t'\delta I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta} f_t(u|x_1,\ldots,x_n)du\right.\right.\\
&\quad \left.\left.-x_t'\delta I_{\{x_t'\delta\le 0\}}\int_{x_t'\delta}^0 f_t(u|x_1,\ldots,x_n)du\right)\right] \qquad \text{(A.13)}\\
&\ge E\left\{\frac{2}{n}\sum I_{\{|x_t'\delta|>\tau\}}\left[x_t'\delta I_{\{x_t'\delta>0\}}\int_0^{x_t'\delta} f_t(u|x_1,\ldots,x_n)du\right.\right.\\
&\quad - \left.\left. x_t'\delta I_{\{x_t'\delta\le 0\}}\int_{x_t'\delta}^0 f_t(u|x_1,\ldots,x_n)du\right][f_t(0|x_1,\ldots,x_n)>f_L]p_1\right\} \qquad \text{(A.14)}\\
&\ge p_1E\left\{\frac{2}{n}\sum I_{\{|x_t'\delta|>\tau\}}\tau f_L d|f_t(0|x_1,\ldots,x_n)>f_L\right\}, \qquad \text{(A.15)}\\
&\ge \tau p_1 f_L d\frac{2}{n}\sum P[|x_t'\delta|>\tau|f_t(0|x_1,\ldots,x_n)>f_L)]. \qquad \text{(A.16)}
\end{aligned}$$

To obtain inequation (A.13), just remark that each term is positive. For the inequation (A.14) we use condition A7. For inequation (A.15) we minorate $|x_t'\delta|$ by τ and each integrals by $f_L d_1$ where $d_1 = \min(\tau, d/2)$. Condition A8 enables us to conclude, by taking the limit,

$$\lim_{n\to\infty} |E[A(\delta)]| \geq 2\tau p_1 f_L d \times \liminf_{n\to\infty} P[|x_i'\delta| > \tau | f_i(0|x_1,\ldots,x_n) > f_L] > 0, \quad \forall \delta > 0 \tag{A.17}$$

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

A.3. Proof of Proposition 5.3

Consider $\hat{\beta}(y, X, u)$ a solution of problem (4.4), let β_0 be the true value of the parameter β and suppose that $u \sim -u$. Equation (5.14) implies that

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

Hence, conditional on X, we have

$$u \sim -u \Rightarrow \hat{\beta}(u, X, u) \sim -\hat{\beta}(-u, X, u) \Rightarrow \mathrm{Med}(\hat{\beta}(\mathrm{u}, \mathrm{X}, \mathrm{u})) = 0. \tag{A.18}$$

Moreover, equation (5.15) implies that

$$\begin{aligned} \hat{\beta}(y, X, u) &= \hat{\beta}(y - X\beta_0, X, u) + \beta_0 \\ &= \hat{\beta}(u, X, u) + \beta_0. \end{aligned} \tag{A.19}$$

Finally, (A.18) and (A.19) imply

$$\mathrm{Med}(\hat{\beta}(\mathrm{y}, \mathrm{X}, \mathrm{u}) - \beta_0) = 0.$$

A.4. Proof of Theorem 5.4

We prove Theorem **5.4** on asymptotic normality. We consider the sign-based estimator $\hat{\beta}(\Omega_n)$ where Ω_n stands for any $p \times p$ positive definite matrix. We apply Theorem 7.2 of Newey and McFadden (1994), which allows to deal with noncontinuous and nondifferentiable objective functions for finite n. Thus, we stand out from usual proofs of asymptotic normality for the LAD or the quantile estimators, for which the objective function is at least continuous. In our case, only the limit objective function is continuous (see the consistency proof). The proof is separated in two parts. First, we show that $L(\beta)$

as defined in equation (5.17) is the derivative of $\lim_{n\to\infty}\frac{1}{n}\sum_t E\big[s\big(u_t - x_t'(\beta-\beta_0)\big)x_t\big]$. Then, we check the conditions for applying Theorem 7.2 of Newey-McFadden.

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

$$\frac{1}{n}\sum_{t=0}^{n} E\big[s\big(u_t - x_t'(\beta-\beta_0)\big)x_t\big] \tag{A.20}$$

is continuous on Θ uniformly over n. Moreover condition A5 specifies that X is $L^{2+\delta}$ bounded. As the $f_t(\lambda|x_1,\ldots,x_n)$ are bounded by f_U uniformly over n and λ (condition A11), dominated convergence allows us to write that

$$\frac{\partial}{\partial\beta}E\big[x_t s\big(u_t - x_t'(\beta-\beta_0)\big)\big] = E\big[x_t x_t' f_t\big(x_t'(\beta-\beta_0)|x_1,\ldots,x_n\big)\big]. \tag{A.21}$$

And, these conditions imply that

$$L_n(\beta) = \frac{1}{n}\sum_{t=1}^{n} E\big[x_t x_t' f_t\big(x_t'(\beta-\beta_0)|x_1,\ldots,x_n\big)\big] \tag{A.22}$$

converges uniformly in β to $L(\beta)$. Uniform convergence entails that $\lim_n \frac{1}{n}\sum_{t=0}^{n} E\big[s\big(u_t - x_t'(\beta-\beta_0)\big)x_t\big]$ is differentiable with derivative $L(\beta)$.

We now apply Theorem 7.2 of Newey and McFadden (1994) which presents asymptotic normality of a minimum distance consistent estimator with nonsmooth objective function and weight matrix $\Omega_n \xrightarrow{p} \Omega$ symmetric positive definite. Thus, under conditions for consistency (A1, A4-A9), we have to check that the following conditions hold:

(i) zero is attained at the limit by β_0;

(ii) the limiting objective function is differentiable at β_0 with derivative $L(\beta_0)$ such that $L(\beta_0)\Omega L(\beta_0)'$ is nonsingular;

(iii) β_0 is an interior point of Θ;

(iv) $\sqrt{n}Q_n(\beta_0) \to \mathcal{N}(0,J)$

(v) for any $\delta_n \to 0$, $\sup_{||\beta-\beta_0||}\sqrt{n}||Q_n(\beta)-Q_n(\beta_0)-EQ(\beta)||/(1+\sqrt{n}||\beta-\beta_0||) \xrightarrow{p} 0$

Condition (i) is fulfilled by the moment condition A1. Condition (ii) is fulfilled by the first part of our proof and condition A13. Then, Condition (iii) is implied by A6. Using the mixing specification A12 of $\{u_t, X_t\}_{t=1,2,\ldots}$ and conditions A1, A5, A10 and A14, we apply a White-Domowitz central limit theorem [see White (2001), Theorem 5.20]. This fulfills condition (iv) of Theorem 7.2 in Newey and McFadden (1994):

$$\sqrt{n}J_n^{-1/2}Q_n(\beta_0) \to N(0, I_p) \tag{A.23}$$

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

$$\sqrt{n}S_n^{-1/2}\big(\hat{\beta}(\Omega_n) - \beta_0\big) \to \mathcal{N}(0, I_p).$$

The asymptotic covariance matrix S is given by the limit of

$$S_n = [L_n(\beta_0)\Omega_n(\beta_0)L_n(\beta_0)]^{-1}L_n(\beta_0)\Omega_n(\beta_0)J_n\Omega_n(\beta_0)L_n(\beta_0)[L_n(\beta_0)\Omega_n(\beta_0)L_n(\beta_0)]^{-1}.$$

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

$$\sqrt{n}S_n^{-1/2}\big(\hat{\beta}(\hat{J}_n^{-1}) - \beta_0\big) \to \mathcal{N}(0, I_p)$$

with

$$S_n = [L_n(\beta_0)\hat{J}_n^{-1}L_n(\beta_0)]^{-1}.$$

When the mediangale Assumption (A2) holds, we find usual results on sign-based estimators. $\hat{\beta}(I_p)$ and $\hat{\beta}[(X'X)^{-1}]$ are asymptotically normal with asymptotic covariance matrix

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

B. Detailed empirical results

This appendix contains additional results for the Barro and Sala-i-Martin application. First, a residual analysis which includes outlier detection, heteroskedasticity tests, etc. is summarized in Table 16. Second, graphics of concentrated sign-based statistics and projected p-values for the β parameter are presented in Figures 27-29.

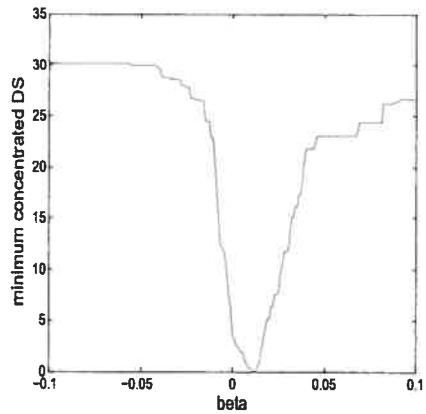
B.1. Regression diagnostics

Table 16. Summary of regression diagnostics.

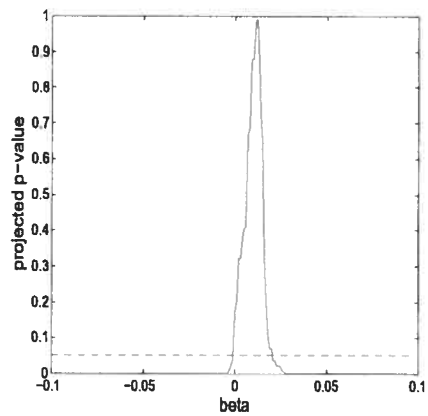
Period	Heterosked.*		Nonnormality**		Influent. obs.**		Possible outliers**	
	Basic eq.	*Eq Reg. Dum.*						
1880-1900	yes	-	yes	-	yes	yes	no	no
1900-1920	yes	yes	yes	yes	yes	yes	yes (MT)	yes
1920-1930	-	-	-	-	yes	-	no	no
1930-1940	-	-	yes	-	yes	yes	no	no
1940-1950	-	-	-	-	yes	yes	yes (VT)	yes (VT)
1950-1960	-	-	-	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 Breush-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, studendized or standardized residuals > 3, DF-beta and Cooks distance have been performed and lead to suspicions for nonnormality, outlier or high influential observation presence.

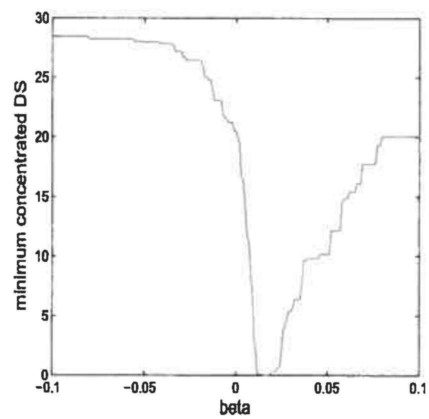
B.2. Concentrated statistics and projected p-values



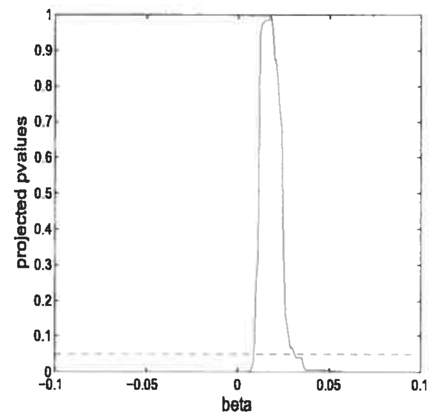
(a) Basic equation: 1880-1900: concentrated DS



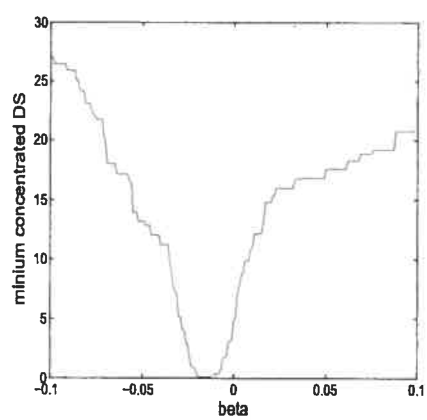
(b) Basic equation: 1880-1900: projected p-value



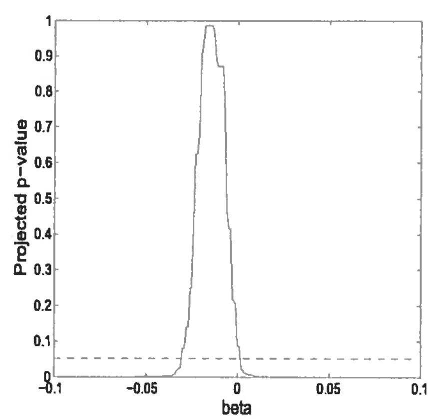
(c) Basic equation: 1900-20: concentrated DS



(d) Basic equation: 1900-20: projected p-value

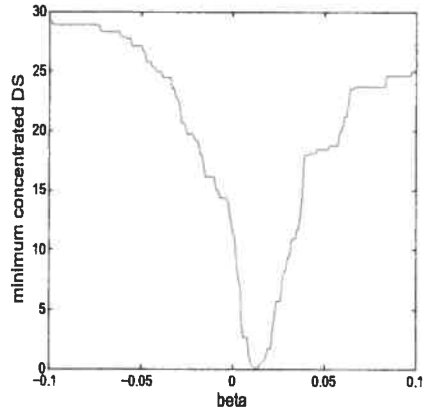


(e) Basic equation: 1920-30: concentrated DS

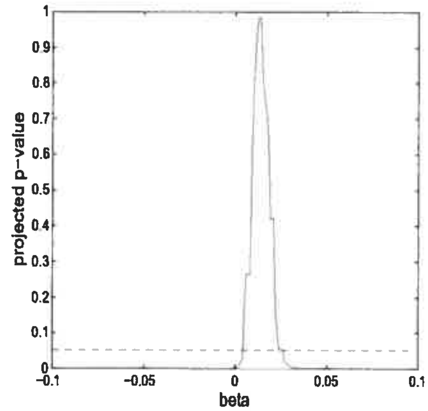


(f) Basic equation: 1920-30: projected p-value

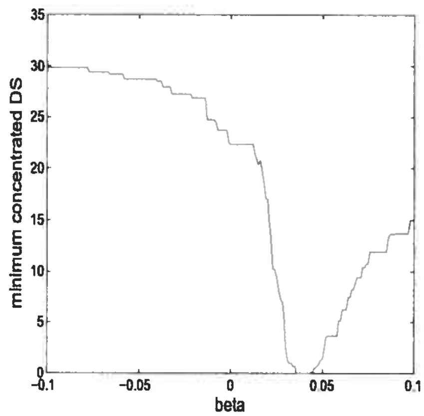
Figure 27. Concentrated statistics and projected p-values (1880-1930)



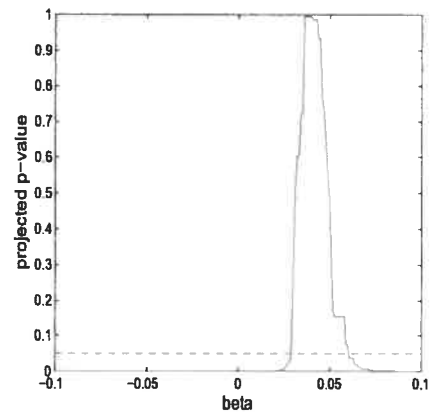
(a) Basic equation: 1930-40: concentrated DS



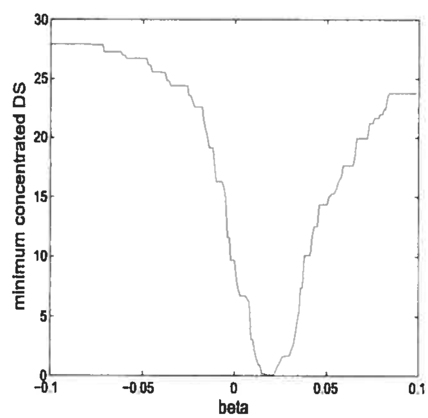
(b) Basic equation: 1930-40: projected p-value



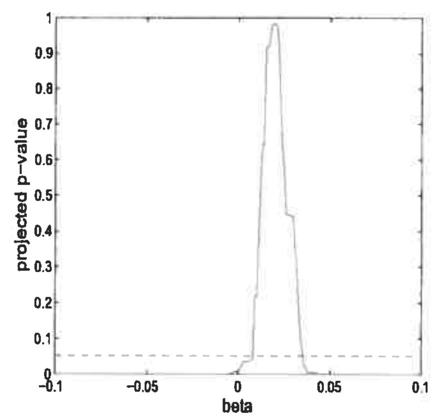
(c) Basic equation: 1940-50: concentrated DS



(d) Basic equation: 1940-50: projected p-value

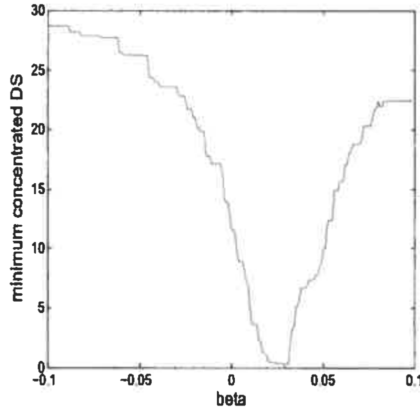


(e) Basic equation: 1950-60: concentrated DS

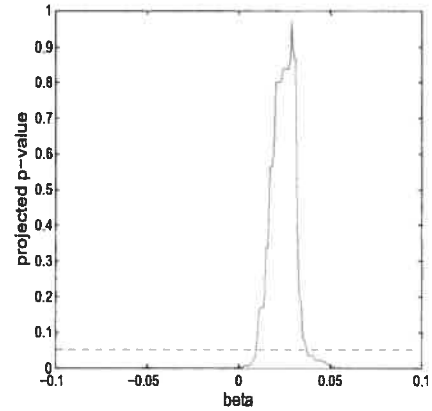


(f) Basic equation: 1950-60: projected p-value

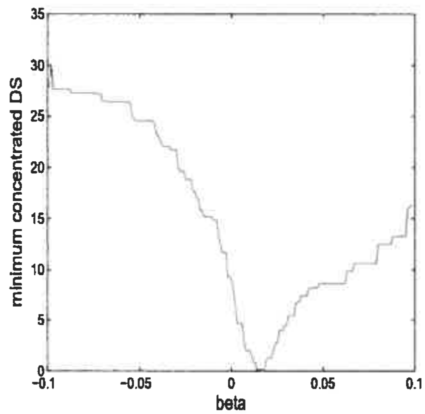
Figure 28. Concentrated statistics and projected p-values (1880-1930)



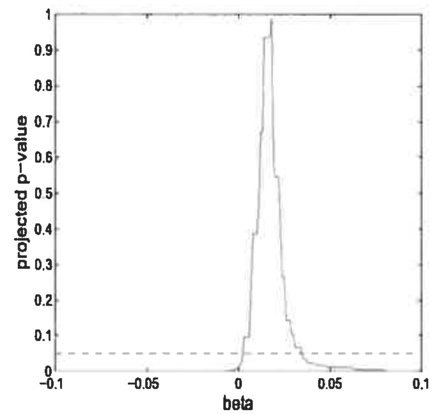
(a) Basic equation: 1960-70: concentrated DS



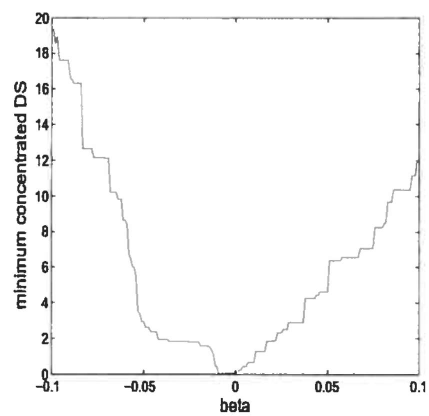
(b) Basic equation: 1960-70: projected p-value



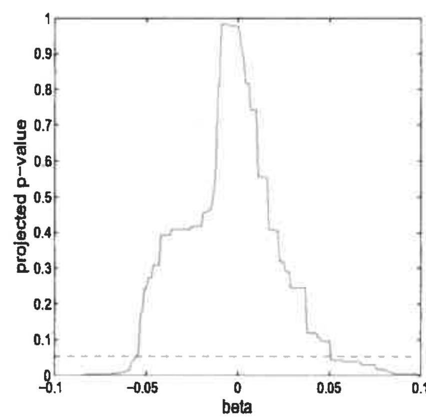
(c) Basic equation: 1970-80: concentrated DS



(d) Basic equation: 1970-80: projected p-value



(e) Basic equation: 1980-88: concentrated DS



(f) Basic equation: 1980-88: projected p-value

Figure 29. Concentrated statistics and projected p-values (1880-1930)

Chapitre 3

Finite and large-sample distribution-free inference in median regressions with instrumental variables

1. Introduction

Instrumental Variable (IV) regression results greatly rely on the quality of the instruments used. When the latter are weakly correlated with the endogenous variable, usual estimators are biased and asymptotic approximations are not anymore valid; see Bound, Jaeger, and Baker (1995), Staiger and Stock (1997), Dufour (1997, 2003), Wang and Zivot (1998), Stock and Wright (2000). Inference relying on estimator asymptotic behavior such as Wald tests may be greatly misleading. One approach to circumvent the problem of weak instruments is to dissociate testing from estimation and to investigate alternative test procedures. Contrary to Wald tests, tests based on the Anderson-Rubin (AR) statistic have correct size for normally distributed disturbances without requiring the parameter to be identified. AR tests are valid in the presence of weak instruments; see Anderson and Rubin (1949), Dufour (1997), Nelson, Startz, and Zivot (1998). However, the AR procedure relies on a Gaussian assumption or at least on some asymptotic justification. In small samples with non-Gaussian disturbances, AR tests (such as any asymptotic test) may be affected by size distortions. Fully exact inference procedures in models where some regressors are endogenous have been less studied. In a regression setup, we propose to use the residual signs to conduct nonparametric valid tests with controlled level for any sample size.

We consider here a possibly nonlinear equation which involves endogenous regressors. A set of exogenous variables is available and no parametric assumption is imposed on the disturbance process. The latter is only assumed to have median zero conditional on the exogenous variables (hereafter, the instruments) and its own past. Without any further restriction, we notice that the sign vector distribution of the constrained residuals is a pivotal function. Its distribution does not depend on nuisance parameters and can easily be simulated. Basically, we use Monte Carlo test techniques [see Dwass (1957), Barnard (1963) and Dufour (2006)] to construct joint sign-based tests that control the level for any sample size. The validity of these tests does not depend on identification assumptions nor on any parametric approximation. In the presence of weak instruments or identification failures, sign-based test levels still equal their nominal size. Then, a complete system of finite-sample inference - as well as asymptotic extensions - can be applied [see Coudin and

Dufour (2005a)]. Simultaneous confidence sets for the whole parameter are obtained by test inversion. Next, confidence sets and tests of general hypotheses are built using projection techniques [see Dufour and Kiviet (1998), Dufour and Jasiak (2001), Dufour and Taamouti (2005)]. Finally, Hodges-Lehmann estimators are provided in identified cases [Hodges and Lehmann (1963), Coudin and Dufour (2005b)]. They correspond to the parameter value least rejected by the tests. As entailed by the results in Dufour (1997), the derived confidence regions may have a non-zero probability of being unbounded in the presence of identification failures.

Nonparametric approaches investigated up to now in the literature have been based on rank and permutation tests. A rank-version of the AR test was introduced by Andrews and Marmer (2005). It dominates the usual AR in terms of size and power for asymmetric and thick tail error distributions. It yields exact tests if the exogenous regressors are independent of instruments and errors. Besides, Bekker and Lawford (2005) proposed exact inference based on permutation tests. Both methods are especially adapted to cross-sectional data, since the errors are assumed to be independent and identically distributed (*i.i.d.*). By contrast, sign-based methods are known to be the only way of producing inference procedures that are proved to be valid under heteroskedasticity of unknown form for a given sample size; see Lehmann and Stein (1949) and Coudin and Dufour (2005a). Sign-based methods provide valid results under very few assumptions. Especially, they allow for general forms of nonlinear dependence in the data. For example, the shape of the error distribution may depend on the instruments provided a sign invariance condition is satisfied. Our approach, which can be applied in time series and in cross-section contexts, extends that part of the literature.

Other test procedures, which are valid in the presence of weak instruments, are parametric or asymptotically justified. A first approach exploits AR-type statistics; see Dufour (1997), Dufour and Jasiak (2001) and Stock and Wright (2000). More recently, Dufour and Taamouti (2005) extended the AR procedure to construct a whole system of inference on the structural parameters. They derived closed-form solutions for the simultaneous confidence regions and for projection-based confidence intervals in special cases. The second approach, followed by Kleibergen (2002, 2005, forthcoming), considered a score-type statistic in the limited information simultaneous equation model (LISEM). The so-called K statistic, which is asymptotically a pivotal function, does not depend on the number of instruments, in contrast with AR tests which loose power when many instruments are involved in the model. In a Gaussian context, Bekker and Kleibergen (2001) investigated the K statistic properties in finite samples. They derived a conservative inference by bounding its behavior. Finally, the conditional approach proposed by Moreira (2003) relies on similar tests; see also Moreira (2001), Moreira and Poi (2003), Cruz and Moreira (2005), Andrews, Moreira, and Stock (2004, 2006). Under the null hypothesis, the size of similar tests does not depend on unknown parameters (especially the endogenous explanatory variables and the instruments). Consequently, a similar test remains valid in the presence of weak instruments. Moreira showed that similar tests can be constructed from non-similar ones by associating a critical value function of those unknown parameters. The conditional likelihood ratio test (CLR) so derived exhibits the best properties. Heteroskedastic and autocorrelation corrected versions of the K and the LR statistics are proposed by Kleibergen (forthcoming). See also Andrews and Stock (2005) for a complete review of the IV literature.

The sign-based approach is in the spirit of Anderson and Rubin.[1] Basically, test statistics are obtained by regressing the signs of the constrained residuals on auxiliary regressors (the instruments) with the particularity that tests are performed using the *exact* distribution of those statistics. Like the AR procedure, a sign-based test may suffer from underrejection when many instruments are involved. This well-known drawback of AR-type procedures is corrected by considering "optimal" instruments which maximize test power. Two optimality concepts are considered: the first one leads to locally optimal tests in the neighborhood of the tested value; the second one to point-optimal tests against a particular alternative. Approximate optimal instruments are constructed by split-sample methods; see Angrist and Krueger (1995), Dufour and Jasiak (2001), Dufour and Taamouti (2005).

Other works on median (and quantile) regression with endogenous regressors have focussed on estimation. The starting point was the two-stage-least-absolute-deviation estimator (2SLAD) introduced by Amemiya (1982), which is an adaptation of 2SLS to the

[1] It is also related to Moreira's approach since the derived tests are similar.

least absolute value (LAV) regression [see also Powell (1983) for the asymptotic properties]. In a first stage, the endogenous variable is regressed by ordinary least squares on the instruments. The second stage consists in a LAV regression which involves the fitted values of the endogenous variable. Chen and Portnoy (1996) extended the idea of two step-estimation to other quantiles. Two robust IV quantile estimators based on GMM formulations are due to Honore and Hu (2004). The first one involves signs of the residuals and the second one their ranks. In a linear median regression model, Hong and Tamer (2003) proposed a minimum distance kernel-based estimator that can be used both in a point identified setup or when there exists a set of observationally equivalent parameters. Besides, control function approaches were used by Lee (2003) in a partially linear quantile regression, by Chernozhukov and Hansen (2004) with a double simultaneous optimization,[2] and by Sakata (2001) who proposed a general approach also based on a double optimization of the ratio between the error dispersion controlled by the instruments and the dispersion without control. Here, we propose to associate a Hodges-Lehmann-type estimator to the finite-sample-based inference results when the parameter is identified. The estimate (or the set of estimates) is the (set of) value(s) least rejected by sign-based tests, or equivalently the one(s) leading to the highest p-value [see Hodges and Lehmann (1963) and Coudin and Dufour (2005b)].

The paper is organized as follows. The model and notations are presented in section 2. In section 3, general results on the finite-sample sign-based inference are stated: the distribution of the constrained signs is derived under the sign invariance assumption. Then, simultaneous tests with controlled level are constructed by Monte Carlo test techniques. Further, confidence sets and general tests are built using projection techniques. In sections 4 and 5, we go further in details and choose the form of the sign-based test statistics on the basis of power properties. Pointwise and local optimality concepts are both considered for choosing the instruments. We also follow two different approaches for determining the form of the sign-based statistic. First, we study a classical GMM statistic that is a quadratic form of the residual signs with a certain weight matrix. We also consider a Tippett-type

[2] Their estimate of the parameter suffering from endogeneity both satisfies the regression criterion minimization and minimizes the instrumental regressors parameters norm. They also obtain valid confidence regions by test inversion.

combination [Tippett (1931)], which relies on the minimum of the p-values corresponding to each sign-based moment equation tested separately. Section 6 is dedicated to asymptotic properties of the proposed test procedures under assumptions weaker than the ones required for finite-sample validity. Section 7 presents IV sign-based estimators when identification holds. The power performances of the sign-based methods are compared to other usual methods in the simulation studies of section 6. Finally, an illustrative application to the returns to schooling [Angrist and Krueger (1991)] is provided in section 8. We conclude in section 9. Appendix A contains the proofs.

2. Framework

In this section, we extend the linear median regression framework used in Coudin and Dufour (2005a) and Coudin and Dufour (2005b) to a nonlinear and instrumental setup. Let $\{W_t = (y_t,\ x_t',\ z_t') : \Omega \to \mathbb{R}^{p+k+1}\}_{t=1,\dots,n}$ be a stochastic process defined on a probability space $(\Omega, \mathcal{F}, P)$, and $\{W_t, \mathcal{F}_t\}_{t=1,\dots,n}$ an adapted stochastic sequence where $\mathcal{F}_t$ is a σ-field in Ω such that $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s < t$ and $\sigma(W_1, \dots, W_t) \subset \mathcal{F}_t$. y_t is the real dependent variable, which can take continuous or discrete values, $x_t = (x_{t1}, \dots, x_{tp})'$ is a p-vector of explanatory variables (possibly endogenous) and $z_t = (z_{t1}, \dots, z_{tk})'$ is a k-vector of exogenous variables. We further assume that y_t, x_t and the parameter of interest, $\theta \in \mathbb{R}^q$, are related through a nonlinear function $f : \mathbb{R}^{1+p+q} \to \mathbb{R}$ up to an error term u_t:

$$f(y_t, x_t, \theta) = u_t,\ t = 1, \dots, n.$$

For convenience, we will use the following matrix notation

$$f(y, X, \theta) = u \tag{2.1}$$

where $y = (y_1, \dots, y_n)'$ and $u = (u_1, \dots, u_n)'$ are real n-vectors, $X = (x_1, \dots, x_n)'$ is a $n \times p$ real matrix.

We denote $Z = (z_1, \dots, z_n)'$ the $n \times k$ real matrix of instruments. The terminology of instruments is very general. It covers exogenous random variables but the instruments may also depend on the parameter θ such as a score vector in a nonlinear model. In such a case, we shall denote $Z_\theta := \big(z_1(\theta), \dots, z_n(\theta)\big)'$. Instruments may be strongly or weakly correlated with the endogenous regressors, but they have to be valid in the following sense.

Assumption A1 Z-CONDITIONAL MEDIANGALE. *Let* $\{u_t, \mathcal{F}_t\}_{t=1,2,\dots}$ *be an adapted stochastic sequence and* $\mathcal{F}_t = \sigma(u_1, \dots, u_t, Z)$. *We assume that*

$$P[u_1 > 0|Z] = P[u_1 < 0|Z] = 1/2,$$
$$P[u_t > 0|Z, u_{t-1}, \dots, u_1] = P[u_t < 0|Z, u_{t-1}, \dots, u_1] = 1/2, \textit{ for } t > 1.$$

Assumption A1 is an adaptation of the mediangale concept defined in Coudin and Dufour (2005a) to an instrumental setup. We condition on Z instead of X since some explanatory variables are endogenous. $\{u_t\}_{t=1,\dots,n}$ are not supposed to be *i.i.d.*. The past values of u_t may have an influence on the form of the distribution of the current u_t, provided they do not affect its probability of being positive or negative. This flexible setup covers the standard limited information simultaneous equations model (LISEM) [see Hausman (1983)]:

$$y_t = x_t'\theta + u_t,$$
$$x_t = z_t'\Pi + v_t,$$
$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} \overset{iid}{\sim} \mathcal{N}(0, \Sigma), \text{ for } t = 1, \dots, n,$$
$$(u_t, v_t') \text{ independent of } z_t, \text{ for } t = 1, \dots, n,$$

where y_t is a scalar dependent variable, x_t is a p-vector of explanatory and possibly endogenous variables, z_t is a k-vector of exogenous variables, u_t is the error term of the structural equation , and v_t is the p-vector of disturbances of the instrumental equation. θ is a p-vector of structural parameters and Π is the $k \times p$ matrix of the reduced form parameters. In a standard LISEM, (u_t, v_t') are *i.i.d.* normally distributed and independent of z_t.

Model (2.1) with the Assumption A1 is much more general. Parametric assumptions on the error term distribution are relaxed. The normality restriction is not required neither in finite samples nor asymptotically. Assumption A1 allows for heteroskedasticity of unknown form. Only the median is assumed to be zero (conditional on Z). This leads to three important special cases.

First, the independence assumption between the observations is relaxed. Past realizations of u_t can have an influence on the shape of the current u_t distribution. For example, $u_t,\ t = 1, \ldots, n$, can satisfy the following assumptions:

$$
\begin{gathered}
u_1 = \sigma_1 \varepsilon_1 \,, \\
u_t = \sigma_t(u_1, \, \ldots \, , \, u_{t-1}) \varepsilon_t \,, \qquad \text{for } t = 2, \, \ldots \, , \, n \\
\varepsilon_1, \, \ldots \, , \, \varepsilon_n \text{ are independent with median zero,} \\
\sigma_1 \text{ and } \{\sigma_t(u_1, \ldots \, , \, u_{t-1})\}_{t=2, \ldots, n} \text{ are non-zero with probability one.}
\end{gathered}
\tag{2.2}
$$

This includes in a time series context ARCH(q) with non-Gaussian noise ε_t, where

$$
\sigma_t(u_1, \, \ldots \, , \, u_{t-1})^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \cdots + \alpha_q u_{t-q}^2 . \tag{2.3}
$$

Second, the instruments may have an influence on the shape of the current u_t distribution, provided the probability of being positive or negative is not affected. In finite samples, an instrument affecting the shape of the disturbance distribution, may be the cause of asymptotic test great distortions. Examples can be found in section 8. In such a case, one can exploit Assumption A1 that allows for some nonlinear dependence between Z and u, for any sample size. A large spectrum of heteroskedastic patterns is covered, such as:

$$
u_t = \sigma_t(Z)\, \varepsilon_t, \ t = 1, \ldots, n, \tag{2.4}
$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are $i.i.d.$ conditional on Z. This can be useful when the instrument choice is limited by data availability.

A third interesting case arises when the endogenous variables affect the shape of the structural error distribution. The usual linear specification simplifies calculus and interpretation. However, if the relation is not well captured by linear modeling, the shape of the structural error distribution may be affected. In such a case, asymptotic tests are invalid even in a large sample.

When u_t and z_t are only asymptotically uncorrelated, Assumption A1 may not hold ($e.g.$ due to feedback on the error signs). However, we will see below that sign-based tests are still asymptotically valid.

3. Finite-sample inference with possibly weak instruments

Assumption A1 is the cornerstone of the validity of sign-based inference methods. If the disturbances satisfy a conditional mediangale condition, their signs have a known joint distribution that does not depend on any nuisance parameter (conditional on the instruments). This property holds for any sample size, without imposing additional distributional assumptions. The sign pivotality property was stated in Coudin and Dufour (2005a) for classical median regressions. It was exploited to construct sign-based simultaneous tests with controlled level for any sample size by Monte Carlo test techniques. In that section, we extend that result to nonlinear and possibly instrumental regressions. Then, we follow the same strategy and conduct simultaneous tests. More generally, the whole finite-sample based inference system presented in Coudin and Dufour (2005a, 2005b) applies here. Simultaneous confidence regions with controlled level are constructed by inverting simultaneous tests; and more general confidence sets or tests, by projecting the simultaneous confidence regions. We rapidly present the leading ideas and principles of finite-sample based inference system. For a detailed presentation, the reader is referred to Coudin and Dufour (2005a, 2005b).

3.1. Pivotality

Let us begin with some notations. We define the sign operator $s : \mathbb{R} \to \{-1, 0, 1\}$ as

$$s(a) = \mathbf{1}_{[0,+\infty)}(a) - \mathbf{1}_{(-\infty,0]}(a), \text{ where } \mathbf{1}_A(a) = \begin{cases} 1, & \text{if } a \in A, \\ 0, & \text{if } a \notin A. \end{cases} \tag{3.1}$$

For convenience, the notation will be extended to vectors. Let $u \in \mathbb{R}^n$ and $s(u)$, the n-vector composed by the signs of its components. This enables us to formally state the following proposition:

Proposition 3.1 SIGN DISTRIBUTION. *Under model* (2.1), *suppose the errors* $(u_1, \ldots, u_n)$ *satisfy Assumption A1 conditional on* Z_θ. *Then the variables* $s(u_1), \ldots, s(u_n)$

are i.i.d. conditional on Z_θ according to the distribution

$$\mathsf{P}_\theta[s(u_t) = 1|Z_\theta] = \mathsf{P}_\theta[s(u_t) = -1|Z_\theta] = 1/2\,, \quad t = 1, \ldots, n\,. \tag{3.2}$$

The proofs of the theorems and propositions appear in the Appendix.

From the latter proposition, it follows that the vector of constrained signs

$$s(f(y, X, \theta)) := \big(s(f(y_1, x_1, \theta)), \ldots, s(f(y_n, x_n, \theta))\big)' \tag{3.3}$$

has a nuisance-parameter-free distribution (conditional on Z), *i.e.* it is a **pivotal function**. When the disturbance process satisfies Assumption A1, the error signs are mutually independent according to a known distribution.

Furthermore, any real-valued function of the form

$$\overline{T}_\theta(y, \theta) = T\big(s(f(y, X, \theta)),\, Z_\theta, \theta\big) \tag{3.4}$$

has a distribution which does not depend on unknown nuisance parameters. Its conditional distribution given Z_θ can be analytically derived or simulated because the joint distribution of $s(f(y, X, \theta))$ is completely specified by Proposition **3.1**. Consequently, we can construct conditional tests for which size is fully controlled.

Consider the problem of testing

$$H_0(\theta_0): \ \theta = \theta_0 \text{ vs } H_1(\theta_0): \theta \neq \theta_0.$$

Under H_0,

$$T\big(s(f(y, X, \theta_0)), Z_{\theta_0}, \theta_0\big) \sim T(S_n, Z_{\theta_0}, \theta_0) \tag{3.5}$$

where $S_n = (s_1, \ldots, s_n)'$ and $s_1, \ldots, s_n$ are *i.i.d.* Bernoulli random variables conditional on Z_{θ_0} that equal 1 with probability $1/2$ and -1 with probability $1/2$. A test with level α rejects the null hypothesis when

$$T\big(s(f(y, X, \theta_0)), Z_{\theta_0}, \theta_0\big) > c_T(Z_{\theta_0}, \alpha, \theta_0) \tag{3.6}$$

where $c_T(Z_{\theta_0}, \alpha, \theta_0)$ is the $(1-\alpha)$-quantile of the distribution of $T(S_n, Z_{\theta_0}, \theta_0)$ conditional on Z_{θ_0}.

This property is an extension of the one stated in Coudin and Dufour (2005a); see also Dufour (1981), Campbell and Dufour (1991, 1995) and Wright (2000).[3] Here, $T\big(s(f(y, X, \theta_0)), Z_{\theta_0}, \theta_0\big)$ and Z_{θ_0} depend on the tested value θ_0. This property can be adapted to error distributions with a mass at zero by randomly breaking the zeros in a way similar to Coudin and Dufour (2005a).

Furthermore, the sign pivotality result allows one to construct nonparametric tests through Monte Carlo test techniques.

3.2. Monte Carlo tests

Under $H_0(\theta_0)$ and Assumption A1, the conditional distribution of $T_{\theta_0}\big(s(f(y, X, \theta_0)), Z_{\theta_0}\big)$ given Z_{θ_0} is free of nuisance parameters with a known distribution that can be simulated. Those two features are sufficient to apply Monte Carlo test procedures.[4] Given T_{θ_0}, the test proposed in section 2 rejects $H_0(\theta_0)$ when $T_{\theta_0} \geq c$, with c depending on the level. The general idea of Monte Carlo tests is to order the observed statistic with N simulated ones. The Monte Carlo test rejects $H_0(\theta_0)$ when the observed statistic is larger than at least $(1-\alpha)\times N$ simulated replicates. As the distribution of T_{θ_0} is discrete, we need a criterion to order two equal realizations. We shall use the randomized tie-breaking presented in Dufour (2006) and Coudin and Dufour (2005a).

The Monte Carlo test for $H_0(\theta_0)$ can equivalently be conducted with empirical p-values. Let $T^{(0)}_{\theta_0}$ be the "observed" statistic, $(T^{(1)}_{\theta_0}, \ldots, T^{(N)}_{\theta_0})$ be a N-vector of independent replicates drawn from the same distribution as T_{θ_0}, and $(W^{(0)}, \ldots, W^{(N)})$ be a $N+1$-vector of $i.i.d.$ real uniform variables. A Monte Carlo test with level α consists in rejecting the null hypothesis whenever the empirical p-value, denoted $\tilde{p}^{\theta_0}_N(T^{(0)}_{\theta_0})$, is smaller than α with

$$\tilde{p}^{\theta_0}_N(x) = \frac{N\tilde{G}^{\theta_0}_N(x) + 1}{N+1}, \tag{3.7}$$

[3]A similar property is stated in Chernozhukov, Hansen, and Jansson (2006) independently from the previous cited works. They use it to compute finite-sample critical values for tests based on a particular GMM statistic, see equation 5.17. They do not use Monte-Carlo version of those tests and restrict on conditionally independent observations.

[4]See Dwass (1957), Barnard (1963) and Dufour (2006)

where

$$\tilde{G}_N^{\theta_0}(x) = 1 - \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{[0,\infty)}(x - T_{\theta_0}^{(i)}) + \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{[0]}(T_{\theta_0}^{(i)} - x)\mathbf{1}_{[0,\infty)}(W^{(i)} - W^{(0)})$$

is the simulated survival function. If N is such that $\alpha(N+1)$ is an integer

$$P[\tilde{p}_N^{\theta_0}(T_{\theta_0}^{(0)}) \leq \alpha] = \alpha \text{ for } 0 \leq \alpha \leq 1.$$

The Monte Carlo test so obtained has size α for any given sample size T. No identification condition is needed to conduct tests with fully controlled level. The instruments may be poorly informative, the test levels are always controlled provided that the instruments are exogenous in the sense of Assumption A1. We shall see later on that Assumption A1 can be slightly relaxed while maintaining the test levels *asymptotically* controlled.

Those basic joint tests constitute the matrix for a whole nonparametric inference system where simultaneous confidence regions are obtained by test inversion and tests of general hypothesis by projection techniques.

3.3. Confidence sets, projection-based confidence intervals and confidence distributions

We use the simultaneous sign-based tests to build confidence sets for θ with given level. These are obtained in the following way: Monte Carlo sign-based tests for $H_0(\theta_0)$ are performed for any value of $\theta_0 \in \mathbb{R}^q$ (or more reasonably for a grid of values) yielding a p-value $\tilde{p}_N^{\theta_0}(T_{\theta_0}^{(0)})$. This associated p-value reflects the *degree of confidence* one may have in the hypothesis $\theta = \theta_0$ given the realization $T_{\theta_0}^{(0)}$ [see Coudin and Dufour (2005b)]. The simultaneous confidence region with level $1-\alpha$ is composed by the values of θ_0 with p-value higher than α. Next, from this simultaneous confidence set for θ, it is possible to derive confidence intervals for the individual components and to perform tests for general nonlinear hypotheses using projection techniques.[5] In Coudin and Dufour (2005b), we

[5] For examples in different settings and for further discussion on projection techniques, the reader is referred to Coudin and Dufour (2005a), Dufour (1990), Dufour (1997), Wang and Zivot (1998), Dufour and Jasiak (2001), Dufour and Taamouti (2005).

directly applied projection techniques on the simulated p-value function. The projected p-value function associated with the individual component θ_k gives a graphical summary of the inference results on θ_k.

The functions involved here are highly nonlinear and no closed-form analytical solutions can easily be obtained. Practical implementation requires to solve optimization problems under nonlinear constraints. Search programs such as simulated annealing are used [see Goffe, Ferrier, and Rogers (1994) and Press, Teukolsky, Vetterling, and Flannery (2002)].

3.4. Simplifications: restrictions on the parameter space

This approach requires in theory to evaluate the sign-based statistic for any value of the parameter in the parameter space. When the size of the parameter space increases, the search programs rapidly become computationally intensive especially when projection techniques are used. So, any additional piece of information that helps to reduce the size of the parameter space is welcome and must be included as a constraint in the program. First of all, restrictions implied by the economic theory or by the relevance of the model have to be taken into account. If the underlying economic model specifies that a certain coefficient must be less than one (such as an elasticity for example), there is no use to investigate what happens outside.

More generally, a conditional approach is also possible. If one accepts to fix some of the parameter components in a certain subspace, say Θ^c, the approach presented above gives results conditional on θ belonging to Θ^c.

An alternative approach consists in restricting the parameter space to a consistent set estimator. Such confidence-set restricted Monte-Carlo tests are asymptotically valid under some general regularity conditions; see Dufour (2006).

The two following sections are dedicated to the construction of efficient test statistics which satisfy the general form $T_\theta\big(s(f(y, X, \theta)), Z_\theta\big)$ so that the finite-sample inference system can be applied. We consider two approaches. First, we establish the general form of point-optimal tests versus a specified alternative. This theoretical result yields a power frontier for sign-based procedures. However, methods that combine various point-optimal tests to approach the power envelope are not easily tractable in practice. Hence, we turn to a more classical approach and derive locally optimal instruments. We study statistics that involve signs in a quadratic form and a Tippett-type combination although other (less usual) statistics could also be envisaged (*e.g.* linear plus quadratic forms or polynomials at various orders involving signs). The class of quadratic IV-type sign-based statistics provides good competitors when the final aim is estimation.

4. Point-optimal tests

Point-optimal tests are usually derived for parametric models since they rely on the likelihood ratio that follows from the classical Neyman-Pearson lemma. Here, they can be constructed for nonparametric models thanks to the sign transformation. In this section, we present point-optimal tests for signs in a general context and then, in a regression context.

4.1. General point-optimal sign-based result

Point-optimal tests based on signs are derived for a very general nonparametric framework in which signs are independent and heterogeneously distributed according to Bernoulli distributions with parameters $(p_1, ..., p_n)$.

$$P[s_t = 1] = p_t, \quad P[s_t = -1] = 1 - p_t, \quad t = 1, \ldots, n. \tag{4.8}$$

Let us consider the problem of testing

$$H_0 : \ (p_1, ..., p_n)' = (p_{01}, ..., p_{0n})', \tag{4.9}$$

against

$$H_1 : \ (p_1, ..., p_n)' = (p_{11}, ..., p_{1n})'. \tag{4.10}$$

Proposition 4.1 POINT-OPTIMAL SIGN-BASED TEST. *When testing H_0 versus H_1, the most powerful test based on signs rejects H_0 when*

$$\sum_{t=1}^{n} s_t \ln\left(\frac{p_{1t}(1-p_{0t})}{p_{0t}(1-p_{1t})}\right) > c(\alpha, H_1)$$

with $c(\alpha, H_1)$ *depending on the level.*

The proof is a direct application of the Neyman-Pearson lemma [see for example Gouriéroux and Monfort (1995b)]. Point-optimal tests are often derived in parametric setups because they rely on the form of the likelihood function under the null hypothesis and under the alternative. Here, the point-optimal test can be derived in a nonparametric setup thanks to the sign transformation. The main strength of the sign transformation is indeed to get rid of the distributional characteristics of the underlying process. However, one has to choose the alternative hypothesis to specify $\{p_{1t}\}_{t=1,\dots,n}$.

4.2. Point-optimal sign-based tests in a regression framework

We now go back to the regression framework of model (2.1) with Assumption A1. Consider testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, Proposition **4.1** yields the following corollary.

Corollary 4.2 POINT-OPTIMAL SIGN-BASED TEST IN A REGRESSION CONTEXT. *In model* (2.1)*, let* $\{W_t = (y_t,\ x_t',\ z_t')\}_{t=1,\dots,n}$ *be a i.i.d. process and* $\{u_t\}_{t=1,\dots,n}$ *have a common distribution function* G *conditional on* Z *that does not depend on* θ*. Suppose further that the mediangale Assumption A1 holds. Then the most powerful sign-based test of* $H_0 : \theta = \theta_0$ *versus* $H_1 : \theta = \theta_1$ *in the sense of Neyman-Pearson rejects* H_0 *when*

$$\sum_{t=1}^{n} s(u_t) \ln\left(\frac{1 - G(h_t)}{G(h_t)}\right) > c(\alpha, \theta_1) \tag{4.11}$$

where $(h_1, \dots, h_n)' = (f(y_1, x_1, \theta_1) - f(y_1, x_1, \theta_0), \dots, f(y_n, x_n, \theta_1) - f(y_n, x_n, \theta_0))'$*, and* $c(\alpha, \theta_1)$ *depends on the level.*

The point-optimal sign-based test is a linear form of the signs with weights depending on the error distribution and the chosen alternative hypothesis. When the distribution function G is logistic, the statistic simplifies and the optimal weights turn out to be $\{h_t\}_{t=1,\dots,n}$.

Point-optimal sign-based tests are theoretically interesting objects because they bound what can be done with signs and combining them allows one to approach the power envelope. However, a point-optimal test requires first to specify the alternative hypothesis and then to compute the optimal weights $\{p_{1t}\}_{t=1,\dots,n}$ that depend on the error distribution. In a parametric setup, this can be done analytically. But in a nonparametric setup (as here), the error distribution is not fixed and $\{p_{1t}\}_{t=1,\ldots,n}$ are not straightforward to choose. Point-optimal statistic can be approached if one "guesses" the behavior of the error term under the alternative hypothesis. This can be done by split-sample techniques. A first part of the sample is used to approach the error distribution, the other part, to construct the statistic; see Dufour and Taamouti (2006) for an example of use.

However, approaching point-optimal tests and power envelope quickly become computationally intensive. For this reason, we turn in the next section to another optimality concept that does not require to specify the alternative hypothesis and still provides "locally" optimal tests. The so-called locally optimal test statistics turn to be quadratic forms of the constrained signs and of optimal instruments. We also study other combinations (than quadratic) of sign-based moment equations that may present power in weak identified cases.

5. IV sign-based statistics

The easiest way to introduce IV sign-based statistics is to refer to a GMM setup. Signs and instruments that satisfy the mediangale Assumption A1 also satisfy usual moment conditions. GMM statistics exploiting the orthogonality between the error signs and the instruments can be constructed using the analogy principle. More generally, we follow the idea of auxiliary regressions [Anderson and Rubin (1949) and Dufour (2003)] to circumvent the problem of endogeneity; see also the artificial regressions of Davidson and McKinnon (2001). We consider regressions of the constrained signs on "auxiliary" instruments (when present in the model their coefficient must be zero). We consider two approaches. IV sign-based statistics correspond either to F-type statistics for testing that the parameter vector in the previous multivariate regression is zero (denoted GMM-type), either to Tippett-type combination of univariate regressions involving one "auxiliary" regressor at once (denoted Tippett-type). The proposed sign-based statistics are pivotal functions and exact sign-based tests can be built for any sample size regardless of the strength of the instruments. Then, we focus on IV sign-based statistics that yield to the best (local) power considerations and on the corresponding optimal instruments.

5.1. Sign-based moment equations

In a usual LISEM model (with valid instruments), the estimating equations correspond to the orthogonality conditions between z_t and u_t.

$$E[(y_t - x_t\theta)z_{jt}] = 0, \quad \text{for } j = 1, \dots, k, \ t = 1, \dots, n. \tag{5.12}$$

Under Assumption A1, Proposition **3.1** entails that the error signs are $i.i.d.$ conditional on Z and centered. Consequently, in model (2.1), the following "sign-based" moment conditions (where the residuals are replaced by their signs) hold:

$$E\big[s\big(f(y_t, x_t, \theta)\big)z_{jt}\big] = 0, \quad \text{for } j = 1, \dots, k, \ t = 1, \dots, n. \tag{5.13}$$

More generally, Assumption A1 entails

$$E\left\{s\big(f(y_t, x_t, \theta)\big)g_j\big(z_t(\theta), \theta\big)\right\} = 0, \quad \text{for } j = 1, \dots, J, \ t = 1, \dots, n. \tag{5.14}$$

where $\{g_j\}_{j=1,\dots,J}$ are measurable functions of the instruments and θ.[6] If necessary, we shall redefine instruments as $\tilde{z}_{jt}(\theta) = g_j\big(z_t(\theta), \theta\big), \ t = 1, \dots, n, \ j = 1, \dots, J$ but the following applies without any further modification.

In those sign-based moment equations, the parameter of interest is not present in an explicit form but is implicitly involved through a robust transformation by the sign operator. The sign operator gets rid of any nuisance parameter affecting the distribution of the error term and enables one to conduct fully robust tests against heteroskedasticity of unknown form for any sample size.

The analogy principle entails the following sample-based moment equations:

$$\sum_{t=1}^{n} s\big(f(y_t, x_t, \theta)\big)z_{jt} = 0, \ j = 1, \dots, k. \tag{5.15}$$

[6]Hong and Tamer (2003) proposed for example to use kernel functions.

5.2. Combining sign-based moment equations: GMM or multiple tests

These new orthogonality conditions can be exploited for constructing GMM-type statistics. For testing $H_0(\theta_0) : \theta = \theta_0$ versus $H_1(\theta_0) : \theta \neq \theta_0$ in model (2.1), we shall consider test statistics of the following form:

$$D_S(\theta_0, Z, \Omega_n) = s(f(y, X, \theta_0))' Z_{\theta_0} \Omega_n\big(s(f(y, X, \theta_0)), Z_{\theta_0}\big) Z_{\theta_0}' s(f(y, X, \theta_0)) \quad (5.16)$$

where $\Omega_n\big(s(f(y, X, \theta_0)), Z_{\theta_0}\big)$ is a $k \times k$ positive definite weight matrix that may depend on the constrained signs $s(f(y, X, \theta_0))$ under $H_0(\theta_0)$.
The statistic associated with $\Omega_n = (Z_{\theta_0}' Z_{\theta_0})^{-1}$ is given by: [7]

$$D_S\big(\theta_0, Z_{\theta_0}, (Z_{\theta_0}' Z_{\theta_0})^{-1}\big) = s\big(f(y, X, \theta_0)\big)' P(Z_{\theta_0}) s\big(f(y, X, \theta_0)\big) \quad (5.17)$$

where $P_{Z_{\theta_0}} = Z_{\theta_0}(Z_{\theta_0}' Z_{\theta_0})^{-1} Z_{\theta_0}'$. That is the squared norm of the fitted values from the regression of $s(f(y, X, \theta_0))$ on Z_{θ_0}. In other words, $D_S\big(\theta_0, Z_{\theta_0}, (Z_{\theta_0}' Z_{\theta_0})^{-1}\big)$ is a monotonic transformation of the Fisher statistic for testing $\gamma = 0$ in the artificial regression model $s\big(f(y, X, \theta_0)\big) = Z_{\theta_0}\gamma + v$.

Another way to approach the problem of building sign-based statistics is then to consider regressions of the constrained signs on appropriately chosen "instruments":

$$s\big(f(y, X, \theta_0)\big) = Z_{\theta_0}\gamma + v \; . \quad (5.18)$$

Testing $H_0(\theta_0)$ is equivalent to test $\gamma = 0$ in (5.18) where $\tilde{Z}(\theta_0)$ are related to X but excluded from the structural model. $\tilde{Z}(\theta_0)$ are called "auxiliary regressors": when present in the model, their coefficient must be zero. Remark that the unilateral point-optimal test presented in Proposition **4.1** can also be viewed as a t-test obtained by regressing the signs on some appropriate auxiliary instruments (precisely the scores under the alternative). Thus, the set of test-statistics based on auxiliary instruments is very general and includes point-optimal Neyman-Pearson-type statistics among the related t-statistics.

[7] This is the GMM statistic studied by Chernozhukov, Hansen, and Jansson (2006) in their conditionally independent setting.

Fisher and GMM-type statistics are quadratic forms of the moment equations. Other types of combination of sign-based moment equations can be exploited. We can for example follow Tippett (1931) and consider

$$D_S^{Tip}(\theta_0, Z_{\theta_0}) = min(p_1, \dots, p_k) \tag{5.19}$$

where $p_1, \dots, p_k$ are the (empirical) p-values associated with testing $\gamma_i = 0$ in the univariate regression involving one instrument (here $z_{i\theta_0}$) at once:

$$s(f(y, X, \theta_0)) = \gamma_i z_{\theta_0 i}, \; i = 1, \dots, k. \tag{5.20}$$

The idea behind is the following. Statistics based on a quadratic combination of moment equations are specifically adapted for test and estimation when the parameter is well identified because they rely a local optimality concept. However, in weakly identified cases, there is no gain to restrict on statistics that provide power in the vecinity of the true value parameter because those values may be observationally equivalent (due to the lack of identification). In such cases, other combinations of the moment equations such as the Tippett combination may provide better overall properties.

5.3. Artificial regressions

The use of artificial regressions such as (5.18) and (5.20) to circumvent endogeneity has been first proposed by Anderson and Rubin (1949) [see also Dufour (2003), Davidson and McKinnon (2001) who presented artificial regressions in general nonlinear models]. In the linear Gaussian model, they proposed an exact test of $\gamma = 0$ based on a Fisher-type statistic. The derived inference is valid and robust to possibly weak instrument settings [see also Dufour (1997), Staiger and Stock (1997), Dufour and Taamouti (2005)]. However, the procedure power depends on the choice of the instruments. In the LISEM model with exact identification and Gaussian disturbances the AR procedure is optimal, but it may suffer from underrejection when a large number of instruments is involved in the model. With "many instruments", asymptotically justified methods such as Kleibergen's K statistic or Moreira's LM statistic may provide better asymptotic power. However, those statistics are no longer pivots in finite samples and a relying inference without other adjustment may suffer from size distortion even in a Gaussian context.[8]

Here, our objective is double. We propose test statistics that are first pivotal functions for any sample size, under the null hypothesis and with known distribution, in order to conduct exact inference (*i.e.* that satisfy Assumption A1) and that are based on an "optimal" choice of instruments.

5.4. Locally optimal instruments

In case of overidentification, instruments can be selected to improve power consideration. When testing $H_0(\theta_0)$ with level α, the power function of the sign-based statistics $T\big(s(f(y, X, \theta_0)), Z_{\theta_0}\big)$ is:

$$\beta(\theta) = P_\theta\big[T\big(s(f(y, X, \theta_0)), Z_{\theta_0}\big) > c_T(Z_{\theta_0}, \alpha)\big]. \tag{5.21}$$

We search for instruments that "maximize" the power function locally around θ_0 in a just identified setup.[9] Around θ_0, sign-based test power functions follow the behavior of their second derivatives $w.r.t.$ θ, which turn to be quadratic forms of the sign vector. Consequently, we derive the optimal instruments from the weights involved in the latter quadratic forms and derive locally optimal sign-based test statistics. This result is stated in the following proposition. Locally optimal instruments are derived in a setup with $i.i.d.$ observations. In the sequel, all results are conditional on the available set of instruments.

Proposition 5.1 LOCALLY OPTIMAL INSTRUMENTS. *Consider the problem of testing* $H_0 : \theta = \theta_0$, *in model* (2.1) *versus a sequence of alternatives* $H_n : \theta = \theta_n$ *such that* $\theta_n \underset{\theta_n \neq \theta_0}{\longrightarrow} \theta_0$, *and assume that:*

a) (y_t, x_t, z_t), $t = 1, \dots, n$ *are identically and continuously distributed;*

b) f *is continuously differentiable in* θ, *with continuous derivative* $H_t(\theta) = \frac{\partial f(y_t, x_t, \theta)}{\partial \theta'}\Big|_\theta$ *and* $H(\theta)' = (H_0(\theta)', \dots, H_n(\theta)')'$ *for* $t = 1, \dots, n$;

[8] The K statistic distribution depends on nuisance parameters in finite samples. In a Gaussian context, Bekker and Kleibergen (2001) derived bounding distributions and conservative tests.

[9] Another alternative is to compute instruments maximizing the power function against a specified alternative. This strategy has been followed by Dufour and Taamouti (2002) who derived point-optimal AR tests in a Gaussian context.

c) $\exists V(\theta_0)$ *such that*

$$\sup_{\theta \in V(\theta_0)} \left|\left| E\left[\frac{\partial f(y_t, x_t, \theta)}{\partial \theta'} \Big|_{\theta} \right] \right|\right| = \sup_{\theta \in V(\theta_0)} ||E[H_t(\theta)]|| \leq M_1, \quad \forall t = 1, \ldots, n;$$

d) u_t *has continuous distribution function* G *which is continuously differentiable at zero with derivative* G' *also continuously differentiable at zero and* $G''(0) = 0$, *for* $t = 1, \ldots, n$;

e) *setting* $P_{\theta_n}\left[u_t - \left(H_t(\overline{\theta}) - EH_t(\overline{\theta})\right)(\theta_n - \theta_0) \leq x\right] = G_n^{\overline{\theta}}(x)$,

$$\tfrac{1}{||\theta_n - \theta_0||}\left(G_n^{\overline{\theta}}(0) - G(0)\right) \to 0 \text{ and } \left(G_n'^{\overline{\theta}}(0) - G'(0)\right) \to 0,$$
$$\text{for all } \overline{\theta} \text{ such that } ||\theta_0 - \overline{\theta}|| \leq ||\theta_0 - \theta_n||.$$

Then, a locally optimal set of instruments is given by

$$Z^*(\theta_0) = E[H(\theta_0)], \tag{5.22}$$

and a locally optimal GMM sign-based statistic is

$$D_S^*(\theta_0) = s\left(f(y, x, \theta_0)\right)' EH(\theta_0)[EH(\theta_0)' EH(\theta_0)]^{-1} EH(\theta_0)' s\left(f(y, x, \theta_0)\right). \tag{5.23}$$

The regularity conditions **b,c** and **d** insure continuity, differentiability and integrability of f and of its derivatives. Condition **d** states that the errors possess a mode at zero. Further, condition **e** sets the speed of convergence of the distribution functions G_n towards G. Further, if $u_t - [H_t(\overline{\theta}) - EH_t(\overline{\theta})](\theta_n - \theta_0)$ has a symmetric distribution for any value of θ_n then condition **e** holds.

If the matrix $H_t(\theta_0)$ is exogenous it can directly be used. If not, we need an exogenous estimate to ensure inference validity for a given n. This is feasible by splitting the sample into two parts.

5.5. Quasi-optimal instruments and split-sample

When observations are independent, one may resort to split-sample techniques.[10] The principle is the following. The sample is divided into two parts: $(Y_{(1)}, X_{(1)}, Z_{(1)})$ and

[10]The split-sample technique was used by Dufour and Taamouti (2002) in a quite similar context to ours. They search an exogenous estimate of the point-optimal matrix of instruments, which, in a Gaussian context, allow them to construct exact inference based on generalized AR statistics, [see also Angrist and Krueger (1995), Dufour and Jasiak (2001) for other uses and a discussion on the optimal split of the sample].

$(Y_{(2)}, X_{(2)}, Z_{(2)})$. The first part is used to estimate

$$\left.\frac{\partial f(Y_{(1)}, X_{(1)}, \theta)}{\partial \theta'}\right|'_{\theta=\theta_0} = h(Z_{(1)}, \theta_0) + \epsilon, \tag{5.24}$$

yielding an estimate $\hat{h}$. This first stage regression may be linear or not, parametric or not depending on the structural model. A sign-based estimation can also be used.

Then, quasi-optimal instruments are constructed for the second part of the sample, $\tilde{Z}_{(2)} = \hat{h}(Z_{(2)})$ and used as auxiliary regressors in the second step regression:

$$s(f(Y_{(2)}, X_{(2)}, \theta_0)) = \gamma \tilde{Z}_{(2)} + v_{(2)}. \tag{5.25}$$

A test of $H_0(\theta_0)$ is thus based on a GMM sign-based statistic

$$SSS(\theta_0) = s\big(f(Y_{(2)}, X_{(2)}, \theta_0)\big)' \tilde{Z}_{(2)}[\tilde{Z}'_{(2)}\tilde{Z}_{(2)}]^{-1}\tilde{Z}'_{(2)} s\big(f(Y_{(2)}, X_{(2)}, \theta_0)\big). \tag{5.26}$$

The latter statistic does not depend on nuisance parameters under the null hypothesis because $\tilde{Z}_{(2)}$ is exogenous. Consequently, Monte Carlo tests can be used. This point also validates the use of simulation-based statistics such as a Tippett-type statistic

$$TSS(\theta_0) = \min\{p_1, \ldots, p_p\} \tag{5.27}$$

where $p_1, \ldots, p_p$ are the empirical p-values for testing $\gamma_i = 0$ in the univariate regressions of the form

$$s(f(Y_{(2)}, X_{(2)}, \theta_0)) = \gamma_i \tilde{z}_{i(2)},\ i = 1, \ldots, p. \tag{5.28}$$

6. Asymptotic properties

A drawback of the mediangale Assumption A1 is the exclusion of linearly dependent processes even though usual asymptotic inference can still be conducted on them. In Coudin and Dufour (2005a), we pointed out that heteroskedasticity and autocorrelation corrected sign-based statistics are asymptotically pivotal functions when signs and explanatory variables are uncorrelated. We also showed that Monte Carlo testing method remained asymptotically valid under weaker distributional assumptions than usual asymptotic Wald tests. In particular, heavy-tailed distributions including infinite variance disturbances were covered. In this section, we show these results apply to IV sign-based statistics without any major modification. We established them for a general nonlinear instrumental regression. A sign HAC-statistic with a weight matrix directly derived from the asymptotic covariance matrix of the signs and the instruments, say $D_S\big(\theta, Z, \frac{1}{n}\hat{J}_n^{-1}(Z)\big)$, turns out to be asymptotically $\chi^2(k)$ distributed under H_0 where k is the number of instruments used.

6.1. Asymptotic behavior of IV GMM sign-statistics

We consider model (2.1) with the following assumptions.

Assumption A2 MIXING. $\{(x_t', z_t'(\theta_0), u_t)\}_{t=1,2,\ldots,}$ *is* α*-mixing of size* $-r/(r-2)$ *with* $r > 2$.[11]

Assumption A3 MOMENT CONDITION. $\mathsf{E}[s(u_t)z_t(\theta_0)] = 0,\ \forall t = 1, \ldots, n,\ \forall n \in \mathbb{N}.$

Assumption A4 BOUNDEDNESS. $z_t(\theta_0) = \big(z_{1t}(\theta_0), \ldots, z_{pt}(\theta_0)\big)'$ *and* $E|z_{ht}(\theta_0)|^r < \Delta < \infty,\ h = 1, \ldots, k,\ t = 1, \ldots, n,\ \forall n \in \mathbb{N}.$

Assumption A5 NON-SINGULARITY. $J_n^{\theta_0} = \mathrm{var}\left[\frac{1}{\sqrt{n}} \sum_{t=1}^{n} s(u_t) z_t(\theta_0)\right]$ *is uniformly positive definite.*

Assumption A6 CONSISTENT ESTIMATOR OF $J_n^{\theta_0}$. $\Omega_n^{\theta_0}$ *is symmetric positive definite uniformly over* n *and* $\Omega_n^{\theta_0} - \frac{1}{n}(J_n^{\theta_0})^{-1} \xrightarrow{p} 0.$

Then we have the following asymptotic distribution.

Theorem 6.1 ASYMPTOTIC DISTRIBUTION OF STATISTIC SHAC. *In model* (2.1), *with Assumptions A2- A6, we have, under* H_0,

$$D_S(\theta_0, Z_{\theta_0}, \Omega_n^{\theta_0}) \to \chi^2(k).$$

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

Corollary 6.2 *In model* (2.1), *with the mediangale Assumption A1 and Assumption A4. If $Z'Z/n$ is positive definite uniformly over n and converges in probability to a definite positive matrix, we have under H_0,*

$$D_S(\theta_0, Z, (Z'Z)^{-1}) \to \chi^2(k).$$

Theorem **6.1** holds for split-sample statistics with $n_2 \to \infty$ and when Z depends on θ (with Z evaluated at θ_0). The proofs are adaptations of Theorem 6.6 and Corollary 6.7 in Coudin and Dufour (2005a).

The $\chi^2(k)$ distribution is familiar in instrumental and weak instruments settings. The statistic $k \times AR$ is asymptotically $\chi^2(k)$ distributed [see Anderson and Rubin (1949), Staiger and Stock (1997), Dufour and Jasiak (2001), Dufour and Taamouti (2005)]. This distribution also bounds the LR and LM statistics [see Wang and Zivot (1998)]. However, the $\chi^2(k)$ distribution is directly related to the number of instruments and the use of many instruments (k large) may entail a power loss. This pleads for the K-statistic favor [see Kleibergen (2002)] in setups with normally distributed disturbances or for any statistic whose distribution does not depend of the number of instruments used. When the setup involves more general processes like non-normal of heteroskedastic errors, there is no reason why the power of a K test would be higher than the one of a sign-based test in finite samples. Nevertheless, if one is concerned about the "many instruments" curse, let us underline that sign-based statistics with quasi-optimal instruments are asymptotically $\chi^2(p)$ distributed as the K-statistic, with the advantage of also providing exact inference in finite samples. Only the combination of a joint testing approach with valid instruments entails exact inference for any sample size.

6.2. Asymptotic validity of Monte Carlo tests

Let a test statistic be asymptotically free of nuisance parameters under H_0, with asymptotic distribution F. Monte Carlo tests that rely on replicates possessing the *same asymptotic distribution* F will asymptotically control the level. This result entails that Monte Carlo tests presented in the previous sections "do at least as well as" asymptotic methods when the mediangale Assumption A1 is relaxed and replaced by a classical moment condition

(Assumption A3); see Coudin and Dufour (2005a). Moreover, those Monte Carlo tests present two considerable advantages over classical asymptotic methods. First, if mediangale Assumption A1 holds, one is sure that the level of Monte Carlo tests is controlled for any sample size. The second advantage comes from the fact that Monte Carlo tests are constructed with replicates based on the *same* sample size. This differs to a classical Monte Carlo test with replicates constructed from the asymptotic distribution. Simulation studies suggest that such Monte Carlo tests perform an implicit sample-size correction [Coudin and Dufour (2005a)]. Indeed, for a given sample size, the distribution of the sign statistic may be closer to the one of the replicates than to the (common) asymptotic distribution. Although the use of such Monte Carlo tests is asymptotically justified, they can be more reliable in small samples than tests based on asymptotic critical values. Under Assumptions A2- A6, testing

$$H_0(\theta_0) : \theta = \theta_0 \text{ versus } H_1(\theta_0) : \theta \neq \theta_0,$$

with the statistic $D_S(\theta, Z_{\theta_0}, \hat{J}_n^{-1}(Z_{\theta_0}))$ is conducted in the following way:

1. Observe $D_S^{(0)} = D_S(\theta_0, Z_{\theta_0}, \hat{J}_n^{-1}(Z_{\theta_0}))$. Draw N replicates of the sign vector as if the n observations were independent. The n components of the replicates are thus independent and drawn from a $B(1, .5)$ distribution.

2. Construct $(D_S^{(1)}, \quad D_S^{(2)}, \ldots, D_S^{(N)})$, the N *pseudo* replicates of $D_S(\theta_0, Z_{\theta_0}, (Z'_{\theta_0} Z_{\theta_0})^{-1})$ under the null hypothesis. We call them *pseudo* replicates because they are drawn as if observations were independent.

3. Draw $N+1$ independent replicates $(W^{(0)}, \ldots, W^{(N)})$ from a $\mathcal{U}_{[0,1]}$ distribution and form the couple $(D_S^{(j)},\ W^{(j)})$.

4. Compute $\hat{p}_n^{(N)}(\theta_0)$ using (3.7).

5. The confidence region $\{\theta \in \mathbb{R}^p | \hat{p}_n^{(N)}(\theta) \geq \alpha\}$ level is at least $1 - \alpha$. We reject $\mathcal{H}_0$ if $\hat{p}_n^{(N)}(\theta_0) \leq \alpha$.

In contrast with Wald-type tests based on LIML or GMM estimators which require identification, those asymptotic results lead to valid inference whatever the informative power of the instruments is and for any degree of identification. Finally, moments and density on the u_t process may not exist.

7. IV sign-based estimators

In the previous sections, we have presented simultaneous tests, confidence sets and more general tests based on signs. Estimation is the last step to a complete the inference system. IV sign-based estimators are obtained in a way similar to the one used for the sign-based estimators studied in Coudin and Dufour (2005b) in a linear regression without instrument. The estimators maximize the p-value function of the parameter given the form of the IV sign-based statistic and the sample size. They present the *highest confidence degree* based on the chosen IV GMM sign-based statistic. They also turn out (with probability one) to minimize the quadratic function of the signs that is given by the sign-based statistic. Here, we introduce IV sign-based estimators for a general nonlinear possibly instrumental regression. We show, for those general models, that they are consistent with asymptotic normal distribution.[12]

7.1. IV sign-based estimators under point identification

When θ is identified, we can define an IV sign-based estimator as any solution $\hat{\theta}_n(\Omega_n)$ of the problem

$$\min_{\theta \in \mathbb{R}^p} s(f(y, X, \theta))' Z_\theta \Omega_n \left(s\left(f(y, X, \theta) \right), Z_\theta \right) Z_\theta' s(f(y, X, \theta)). \tag{7.29}$$

IV sign-based estimators are analogues of sign-based estimators studied in Coudin and Dufour (2005b). These constitute Hodges-Lehmann-type estimators in the sense that they are associated with the highest degree of confidence one may have in a value of θ given the realization of the sample and the choice of the sign-based test statistic $D_S(Z_\theta, \Omega_n, \theta)$ [Hodges and Lehmann (1963)]. The reader is referred to Coudin and Dufour (2005b) for a detailed presentation. IV-sign based estimators can also be interpreted as GMM estimators

[12] Estimators based on the Tippett-sign statistic could be defined as solutions of a double optimization problem: maximization of the minimal p-value (a sort of Rawls criteria between the moment equations). That question is not addressed in the present paper.

exploiting the orthogonality between error signs and instruments. See Honore and Hu (2004) for a presentation in an instrumental linear regression with $i.i.d.$ disturbances and Coudin and Dufour (2005b) for equivalence (with probability one) between both definitions.

For practical use, we also introduce a two-step estimator $\hat{\theta}_n^{2S}(\Omega_n)$ as any solution of the problem

$$\min_{\theta \in \mathbb{R}^p} s(f(y, X, \theta))' Z_\theta \Omega_n(s(f(Y, X, \hat{\theta}_n)), Z_{\hat{\theta}_n}) Z_\theta' s(f(y, X, \theta)), \tag{7.30}$$

where $\hat{\theta}_n$ is a first stage consistent estimator.

In the following, we show that the IV sign-based estimators defined in equations (7.29) and (7.30) are consistent and asymptotically normal if the parameter is identified.

7.2. Consistency

We first prove the consistency of IV sign-based estimators when the auxiliary regressors are integrable and continuous functions of the parameter θ and of some l-vector process v_t, $t = 1, 2, \ldots$, on which the mixing conditions are imposed. Let $h_t : \Theta \times \mathbb{R}^l \longrightarrow \mathbb{R}^k$, $\forall t$,

$$z_t(\theta) = h_t(\theta, v_t), \; t = 1, \ldots. \tag{7.31}$$

We assume that the following conditions hold.

Assumption A7 MIXING. *$\{W_t^v = (y_t, x_t', v_t')\}_{t=1,2,\ldots}$ is α-mixing of size $-r/(r-1)$ with $r > 1$.*

Assumption A8 CONTINUITY OF F. *$f(y_t, x_t, \theta)$ is measurable, a.e. continuous in θ with $P[f(y_t, x_t, \theta) = 0] = 0$, $\forall \theta \in \Theta$.*

Assumption A9 BOUNDEDNESS AND CONTINUITY.

a) *$z_t(\theta) = (z_{1t}(\theta), \ldots, z_{pt}(\theta))'$ and $E|z_{ht}(\theta)|^{r+1} < \Delta < \infty$, $h = 1, \ldots, k$, $t = 1, \ldots, n$, $\forall n \in \mathbb{N}$, $\forall \theta \in \Theta$.*

b) $z_{ht}(\theta)$ *is a.e. continuous in* $\theta, \forall t$.

c) $P[z_{ht}(\theta) = 0] = 0,\ \forall \theta \in \Theta, \forall t$.

Assumption A10 COMPACTNESS. $\theta \in Int(\Theta)$, *where* Θ *is a compact subset of* $\mathbb{R}^p$.

Assumption A11 POINT IDENTIFICATION.

$$\lim_{n\to\infty} E\left[\frac{1}{n}\sum_t s\big(f(y_t, x_t, \theta)\big) \otimes z_t(\theta)\right] = 0 \Rightarrow \theta = \theta_0$$

Assumption A12 UNIFORMLY POSITIVE DEFINITE WEIGHT MATRIX. $\Omega_n(\theta)$ *is symmetric positive definite for all* θ *in* Θ.

Assumption A13 LOCALLY POSITIVE DEFINITE WEIGHT MATRIX NEAR θ_0. $\Omega_n(\theta)$ *is symmetric positive definite for all* θ *in a neighborhood of* θ_0.

The mixing condition (Assumption A7) is imposed on a underlying process, $\{v_t\}_{t=1,2,\ldots}$, because the instruments are functions of the parameter. Assumptions A8 and A9 contain the regularity conditions required on the functions f and h_t. Remark in particular that the sets of zeros are assumed to be negligible. Assumption A10 is the classical compactness condition. Assumptions A11, A12 and A13 are classical and required for identification. Then we have the following property.

Theorem 7.1 CONSISTENCY. *Under model* (2.1) *with the Assumptions A3 and A7-A12, any IV sign-based estimator defined by* (7.29) *is consistent.*

When Assumption A12 is replaced by Assumption A13, the two-step estimators defined in (7.30) are consistent. Consistency is established without requiring second-order moment existence of the disturbances u_t. Indeed, the disturbances appear in the objective function only through their sign transforms which possess finite moments at any order. Consequently no additional restriction should be imposed on the disturbance process. Those points also entail a more general CLT than usual.

7.3. Asymptotic normality

Asymptotic normality requires some additional assumptions.

Assumption A14 UNIFORMLY BOUNDED DENSITIES. $\exists g_U < +\infty$ *such that* , $\forall n \in \mathbb{N}, \forall \lambda \in \mathbb{R}$,

$$\sup_{1 \leq t \leq n} |g_t(\lambda | x_1, \ldots, x_n)| < g_U, \; a.s.$$

Assumption A15 DIFFERENTIABILITY OF f. *f is a.e. continuously differentiable in θ and* $E||\frac{\partial f}{\partial \theta'}|_\theta|| < +\infty, \; \forall \theta \in \Theta$.

Assumption A16 MIXING WITH $r > 2$. $\{W_t = (y_t, x_t', v_t')\}_{t=1,2,\ldots,}$ *is α-mixing of size* $-r/(r-2)$ *with* $r > 2$.

Assumption A17 DIFFERENTIABILITY OF h. $z_t = h_t(\theta, v_t)$ *and h_t is a.e. continuously differentiable in θ and* $E||\frac{\partial h_t}{\partial \theta'}|_\theta|| < +\infty, \; \forall \theta \in \Theta, \; \forall t = 1, \ldots, n, \; \forall n \in \mathbb{N}$.

Assumption A18 DEFINITE POSITIVENESS OF $J_n(\theta_0)$. *$J_n(\theta_0)$ is $k \times k$ and uniformly positive definite in n and converges to a definite positive symmetric matrix J, where,* $J_n(\theta) = \text{var}\left[\frac{1}{\sqrt{n}} \sum_t^n s(u_t) h_t(\theta, v_t)\right]$.

Assumption A19 DEFINITION OF L_n. *$L_n(\theta_0)$ is a $p \times k$ matrix defined as:*

$$\begin{aligned} L_n(\theta) &= \frac{1}{n} \sum_t E\left[h_t(\theta, v_t) \frac{\partial f(y_t, x_t, \theta)}{\partial \theta'} g_t\big(f(y_t, x_t, \theta) | z_1, \ldots, z_n\big)\right] \\ &+ \frac{1}{n} \sum_t E\left[\frac{\partial h_t}{\partial \theta'} s\big(f(y_t, x_t, \theta)\big)\right]. \end{aligned}$$

$L_n'(\theta_0) \Omega_n L_n(\theta_0)$ is nonsingular uniformly in n.

Assumption A16 is the classical mixing condition required in asymptotic normality proofs. Assumptions A15, A17 and A19 are regularity conditions for nonlinear setups. Assumption A14 is usual in the LAD and quantile theory: bounded variance conditions (horizontal spread) are replaced by bounded vertical spreads. Assumption A18 is classical. We see in Assumption A19 that $L_n(\theta)$ has a second term induced by the fact that the instruments depend on the parameter. Then, we have the following theorem.

Theorem 7.2 ASYMPTOTIC NORMALITY. *Under the conditions for consistency and Assumptions A14-A19 we have:*

$$S_n^{-1/2}\sqrt{n}\big(\hat{\theta}_n(\Omega_n) - \theta_0\big) \xrightarrow{d} N(0, I_p) \tag{7.32}$$

where

$$S_n = \ [L_n(\theta_0)\Omega_n L_n(\theta_0)']^{-1} L_n(\theta_0)\Omega_n J_n \Omega_n L_n(\theta_0)'[L_n(\theta_0)\Omega_n L_n(\theta_0)']^{-1}.$$

When $\Omega_n = \hat{J}_n^{-1}$,

$$[L_n(\theta_0)\hat{J}_n^{-1} L_n(\theta_0)]^{-1/2}\sqrt{n}\big(\hat{\theta}_n(\hat{J}_n^{-1}) - \theta_0\big) \xrightarrow{d} N(0, I_p). \tag{7.33}$$

Theorem 7.2 holds in particular for classical instrumental setups when the instruments Z do not depend on θ. In such a case, $L_n(\theta)$ simplifies to

$$L_n(\theta) = \frac{1}{n}\sum_t E\left[z_t \frac{\partial f(y_t, x_t, \theta)}{\partial \theta'} g_t(f(y_t, x_t, \theta)|z_1, \ldots, z_n)\right]. \tag{7.34}$$

This result extends the classical sign-based estimator asymptotic normality established in Coudin and Dufour (2005b) for nonlinear and instrumental regressions. Note again the existence of the second-order moment disturbances is not required. The sign asymptotic normality holds for heavy-tail distributions whereas usual estimators, such as the 2SLS estimator, do not. The dispersion measure adapted to sign-based estimators do not refer to the error variance but to the (inverse of the) error density evaluated at zero. This alternative dispersion measure, called the "diffusivity", is involved in Cramér-Rao type lower bound for median-unbiased estimators; see Coudin and Dufour (2005b), Sung, Stangenhaus, and David (1990) and So (1994).

The properties of consistency and asymptotic normality entirely rely on the identification assumption whereas the sign-based inference presented previously does not. This provides the occasion to recall the main message of the weak IV literature: when some identification failure or the presence of weak instruments are suspected, tests based on the asymptotic behavior of estimators should be avoided. Inference should be based on test statistics that are robust to identification failure such as IV sign-based statistics. The next section illustrates by a simulation study, how important it can be to use the exact distribution of such robust statistics.

8. Simulation study

In this section, we present simulation studies comparing the performance of sign-based methods with usual instrument-based techniques. We consider the basic sign-based statistic $D_S(\theta, Z, (Z'Z)^{-1})$ (denoted BS) and a split-sample based one that aims to overcome possibly power loss when "many instruments" are used (SSS). We compare tests based on those two statistics with Wald tests based on the 2SLS estimator and the 2SLAD estimator (both estimators are unreliable in the presence of weak instruments), and with some tests that are "robust to weak instruments". Those robust tests rely on the Anderson-Rubin statistic (AR) [Anderson and Rubin (1949)], the Anderson-Rubin statistic with split-sample (SSAR) [Dufour and Jasiak (2001)], the score statistic proposed by Kleibergen (2002) (K) and the score statistic corrected for heteroskedasticity (KLM) [Kleibergen (forthcoming)]. We use the following linear model taken from Kleibergen (2002) with different numbers of instruments, degrees of identification and various disturbance behaviors:

$$y = Y\theta + \epsilon$$
$$Y = X\Pi + V,$$

where n is the number of observations, $y,\ Y\ :\ n \times 1$, $X\ :\ n \times k$, $X \sim \mathcal{N}(0, I_k \otimes I_n)$, $\Pi\ :\ k \times 1$, $\theta = 0$. In $\Pi = (\pi_1, \ldots, \pi_k)'$, four different values of π_1 are considered: 1 (strong valid instrument), 0.5 (instrument of mild strength), 0.1 (weak instrument), and 0 (no identification). Other components of Π are set to zero. The number of instruments k alternatively equals 1, 5 or 10 in view of studying the effect of including irrelevant instruments.

We wonder what the test performances are for various schemes of disturbances. Therefore, we do not restrict on $i.i.d.$ normal disturbances. We also study heavy-tailed disturbances and heteroskedastic schemes. We use the four following data generating processes:

Case 1: *i.i.d.* normal disturbances:

$$(\epsilon, V) \sim \mathcal{N}(0, \Sigma \otimes I_n),\ \Sigma = \begin{pmatrix} 1 & .99 \\ .99 & 1 \end{pmatrix}.$$

Case 2: *i.i.d.* Cauchy disturbances:

$$(\epsilon^1, V^1) \sim \mathcal{C} \text{ and } (\epsilon_t, V_t)' = \Sigma(\epsilon_t^1, V_t^1)', \text{ with } \Sigma = \begin{pmatrix} 1 & .99 \\ .99 & 1 \end{pmatrix}.$$

Case 3: some instruments affect the shape of the structural error ϵ :

$$(\epsilon^1, V) \sim \mathcal{N}(0, \Sigma \otimes I_n),\ \epsilon_t = x_{t1}^2 \epsilon_t^1, t = 1, \ldots, T.$$

Case 4: the endogenous variable affects the shape of ϵ:

$$(\epsilon^1, V) \sim \mathcal{N}(0, \Sigma \otimes I_n),\ \epsilon_t = Y_t^2 \epsilon_t^1, t = 1, \ldots, T.$$

Cases 1 and 2 illustrate the effect of a departure from normality on the different tests: homoskedastic disturbances, which are normally distributed in case 1 and Cauchy distributed in case 2. In normal cases, with one instrument, the K statistic which equals the AR is optimal. We wonder what happens when normality is relaxed and especially when the disturbances possess heavy tails. The next DGPs (cases 3 and 4) illustrate heteroskedasticity. In case 3, the instruments affect the variance of the structural error. In case 4, the endogenous variable affects the variance of the structural error. We illustrate how the classical tests (K, AR) fail in the presence of heteroskedasticity and we focus on comparing sign-based tests to the KLM tests that are corrected for heteroscedasticity. Remark that for the four cases, the mediangale Assumption A1 holds and sign-based methods do exactly control levels for any sample size.

8.1. Size

We first investigate level distortions. We consider the testing problem: $H_0 : \theta_0 = 0$ against $H_1 : \theta_0 \neq 0$, and report empirical rejection frequencies for tests of level .05.

Empirical sizes are computed using 10000 simulations. Bootstrap and Monte Carlo methods are both based on 2999 replicates. For split-sample statistics (SSAR and SSS), 15 observations are used for the first stage and 35 for the second stage.

Sign-based tests (BS, SSS) are the only ones that have perfectly controlled levels in the four presented cases. Empirical sizes of sign-based tests equal the nominal size. In contrast, empirical sizes of Wald tests (2SLAD, 2SLS) greatly suffer from the small number of observations, the weakness of the instruments and the presence of irrelevant instruments. The empirical sizes of the AR, SSAR and K tests are smaller than the Wald-type test ones in homoskedastic setups because their asymptotic levels equal the nominal one whatever the strength and the number of instruments. However, they are affected by finite-sample distortions and loose their relevance in heteroskedastic setups. Finally, tests based on the KLM statistic involving a White-type correction for heteroskedasticity have empirical sizes close to the nominal one for setup 3, but this is no longer true when endogeneity affects the variance of the structural error (setup 4).

Simulations confirm the theory. Sign-based tests allow to control test levels for a very wide range of setups and for any sample size. They are the only ones that are robust to heteroskedasticity of unknown form.

Table 17. Empirical sizes: n=50.

Case 1 : *i.i.d.* normal distribution												
nb inst.	k=1				k=5				k=10			
π_1	1	.5	.1	0	1	.5	.1	0	1	.5	.1	0
W2SLS	.087	.123	.375	.911	.315	.708	.994	1.00	.548	.939	1.00	1.00
W2SLAD	.028	.019	.001	.000	.161	.352	.691	.715	.296	.595	.873	.889
AR	.059	.059	.059	.059	.067	.067	.067	.067	.088	.088	.088	.088
SSAR	.116	.116	.116	.116	.095	.096	.097	.097	.085	.086	.084	.084
K	.059	.059	.059	.059	.057	.057	.056	.070	.060	.060	.060	.088
KLM	.048	.048	.048	.048	.024	.024	.024	.036	.016	.016	.016	.032
BS	.050	.050	.050	.050	.045	.045	.045	.045	.056	.056	.056	.056
SSS	.052	.052	.052	.052	.049	.048	.047	.047	.052	.050	.051	.051
Case 2 : *i.i.d.* Cauchy distribution												
π_1	1	.5	.1	0	1	.5	.1	0	1	.5	.1	0
W2SLS	.477	.607	.822	.937	.987	.998	1.00	1.00	1.00	1.00	1.00	1.00
W2SLAD	.001	.001	.000	.000	.037	.037	.038	.036	.045	.047	.048	.047
AR	.061	.061	.061	.061	.063	.063	.063	.063	.081	.081	.081	.081
SSAR	.121	.121	.121	.121	.103	.103	.102	.102	.080	.082	.081	.081
K	.061	.061	.061	.061	.054	.054	.055	.066	.066	.067	.067	.077
KLM	.019	.019	.019	.019	.034	.034	.034	.032	.027	.028	.028	.029
BS	.051	.051	.051	.051	.053	.053	.053	.053	.056	.056	.056	.056
SSS	.050	.050	.050	.050	.047	.047	.047	.047	.056	.053	.056	.055
Case 3 : instruments affect the shape of error distribution												
π_1	1	.5	.1	0	1	.5	.1	0	1	.5	.1	0
W2SLS	.101	.129	.203	.213	.140	.256	.475	.493	.160	.328	.674	.700
W2SLAD	.021	.015	.004	.003	.048	.039	.017	.016	.088	.081	.047	.044
AR	.417	.417	.417	.417	.249	.249	.249	.249	.223	.223	.223	.223
SSAR	.510	.510	.510	.510	.280	.215	.184	.179	.179	.131	.111	.111
K	.417	.417	.417	.417	.329	.263	.159	.153	.357	.259	.129	.120
KLM	.029	.029	.029	.029	.026	.034	.040	.040	.032	.038	.043	.042
BS	.053	.053	.053	.053	.048	.048	.048	.048	.057	.057	.057	.057
SSS	.053	.053	.053	.053	.055	.051	.052	.050	.051	.051	.053	.054
Case 4 : endogeneity affects the shape of error distribution												
π_1	1	.5	.1	0	1	.5	.1	0	1	.5	.1	0
W2SLS	.744	.519	.234	.216	.898	.849	.821	.822	.923	.967	.972	.972
W2SLAD	.012	.006	.001	.001	.030	.028	.027	.026	.056	.059	.062	.064
AR	.526	.220	.068	.061	.300	.128	.072	.069	.323	.162	.084	.080
SSAR	.527	.269	.128	.121	.282	.135	.097	.096	.221	.108	.081	.079
K	.526	.220	.068	.061	.406	.128	.068	.068	.497	.169	.081	.082
KLM	.321	.126	.032	.028	.207	.077	.040	.039	.055	.068	.044	.041
BS	.051	.051	.051	.051	.044	.044	.044	.044	.054	.054	.054	.054
SSS	.050	.050	.050	.050	.049	.052	.051	.051	.049	.051	.050	.050

8.2. Power

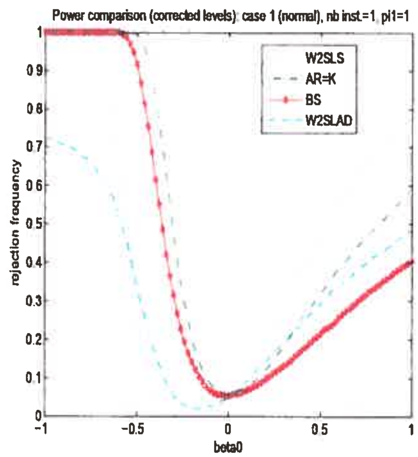
Then, we compare the power of these tests. Tests of $H_0 : \theta = 0$ are performed on data obtained by letting vary θ. Simulated power is given by a graph with θ in abscissa; see Figures 30, 31, 32, 33. The power functions presented here are locally adjusted for the level when needed, which allows comparisons between methods. However, we should keep in mind that only sign-based tests do exactly control the level for any sample size. All results concerning homoskedastic or heteroskedastic setups with a given number of instruments and for various instrument strength are contained in a single figure. In Figures 30 and 31, errors are homoskedastic, either normal (first column), either Cauchy (second column). The number of instruments equals one for Figure 30, and five for Figure 31. Therefore, comparing both columns illustrates which effect a departure from normality (here Cauchy disturbances) entails on the test powers. The effect of heteroskedasticity is then illustrated by Figures 32 (model with one instrument) and 33 (model with five instruments). We are particularly interested in comparing the sign-based method to the KLM method (and 2SLAD, 2SLS for strong instruments) which is corrected for heteroskedasticity since the K and the AR methods are not.

Let us now examine the results. In a model with one instrument (Figure 30), the K statistic and the AR statistic are equal. The AR statistic is best for the $i.i.d.$ normal case 1 but the sign-based power curve is not far from that optimal power curve (first column of Figure 30). With Cauchy distributions (case 2, column 2 in Figure30), the sign-based power curve is far above all the others. This holds regardless of instrument strength. The power curves of Wald tests based on the 2SLS and the 2SLAD estimators are also reported when the instruments are strong. In case 1, these methods are biased; in case 2, they do not present power anymore.

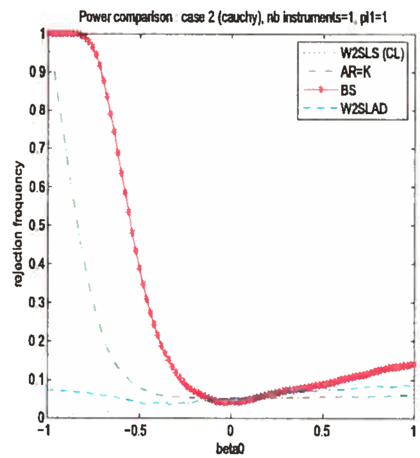
The AR procedure and the sign-based procedure loose power as the number of (irrelevant) instruments included in the model increases. Figure 31 illustrates the power curves when the model involves five instruments. For the $i.i.d.$ normal case (case 1, column 1 in Figure31), the K statistic, which now differs from the AR statistic, does not encounter this loss of power and leads to the highest power curve whereas both the sign-based power curve and the AR-based one stand lower. However, as soon as we turn to the Cauchy setup (case 2, column 2 in Figure31), the sign-based statistic yields again the highest power. This holds regardless of instrument strength. The two methods involving a split-sample (SSAR and SSS) do not present good results because of the limited number of observations. Here, the sample size is 50. First step regressions involve only 15 observations and second step regressions 35 observations. However, the corresponding power curves generally follow the same tendencies as the power curves of the corresponding statistic *without* split-sample.

Results are very clear in Figures 32 and 33 (heteroskedastic setups: case 3 and 4). Sign-based methods exhibit there more power than all the other studied methods which are robust to weak instruments (AR, K) included methods corrected for heteroskedasticity (KLM). In the presence of strong instruments, Wald tests based on 2SLAD and 2SLS have higher power than sign-based methods. However, the Wald tests are clearly biased and they are no longer valid as soon as the strength of the instruments decreases.

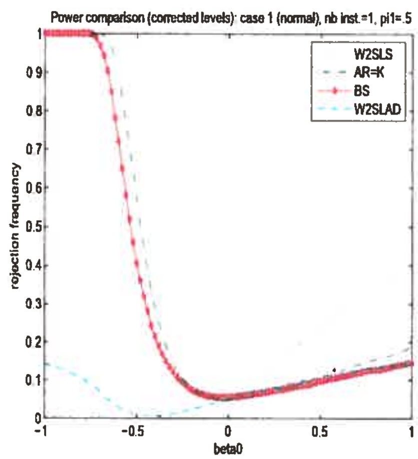
In conclusion, sign-based tests present good power properties for a wide range of processes. They are not far from the optimal AR test in $i.i.d.$ normal case and they provide more power than other studied methods in setups involving heavy-tailed distributions, heteroskedasticity or nonlinear dependence. They still provide power under some general endogeneity schemes, especially when the endogeneity affects the shape of the structural error distribution without affecting its sign.



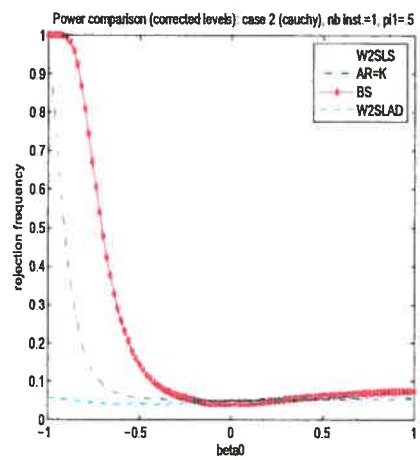
(a) Case 1 strong instrument



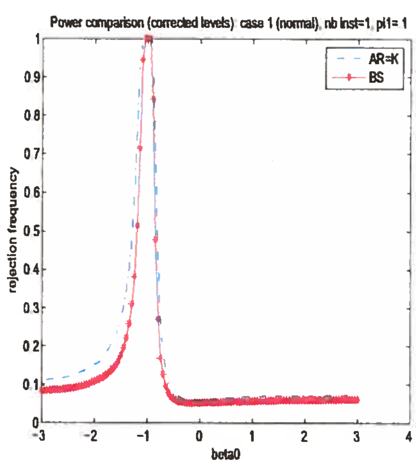
(b) Case 2 strong instrument



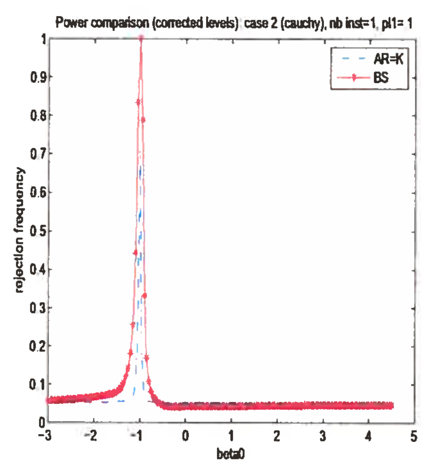
(c) Case 1 instrument .5



(d) Case 2 instrument .5

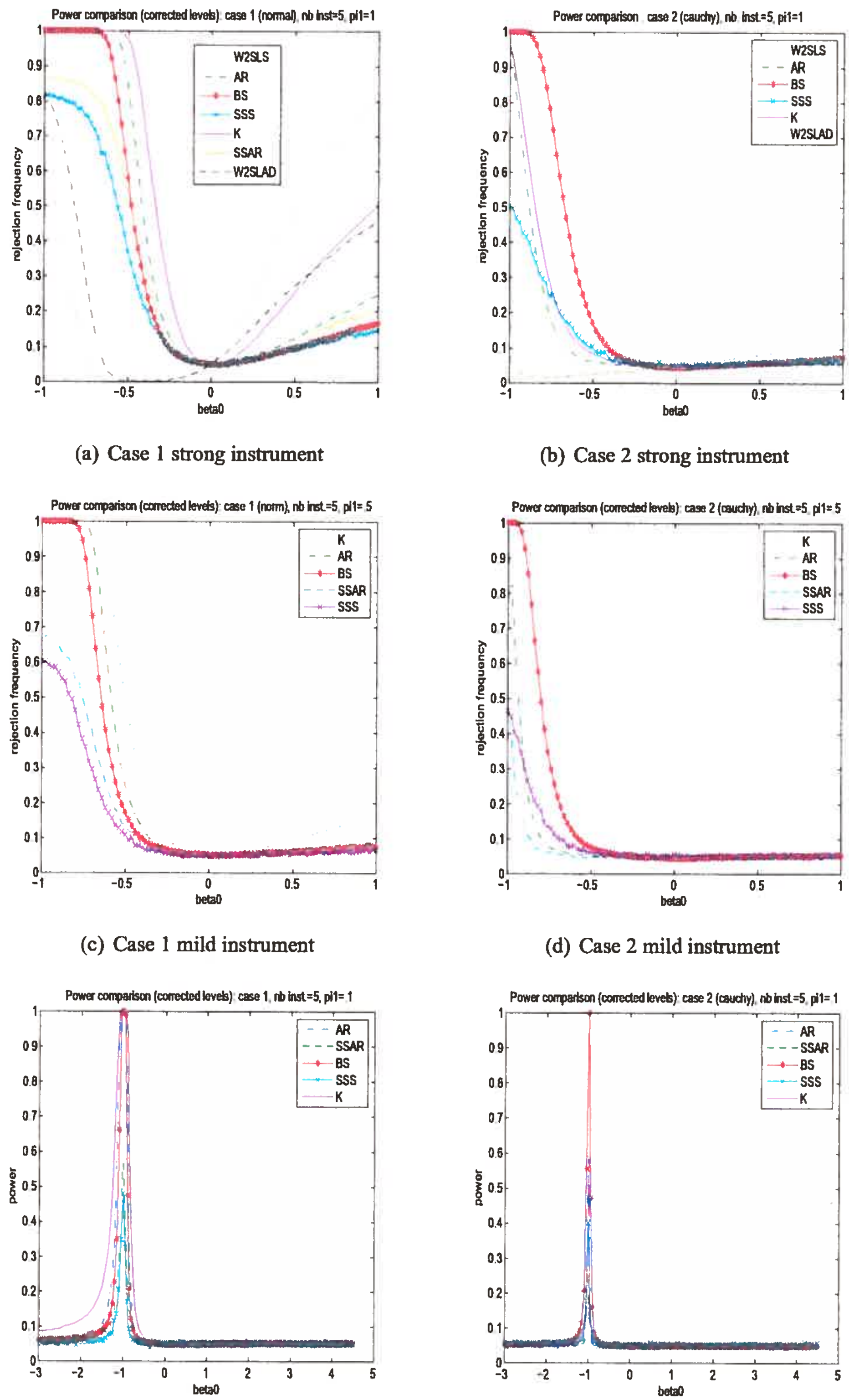


(e) Case 1 weak instrument



(f) Case 2 weak instrument

Figure 30. Power functions: model with one instrument.



(a) Case 1 strong instrument

(b) Case 2 strong instrument

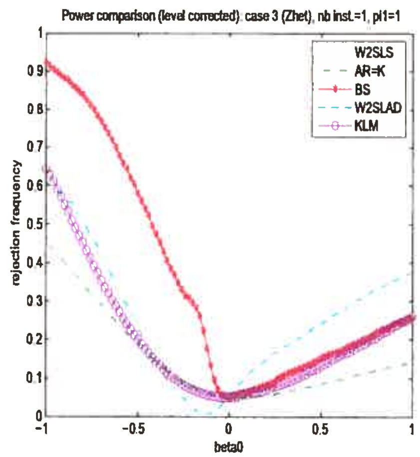
(c) Case 1 mild instrument

(d) Case 2 mild instrument

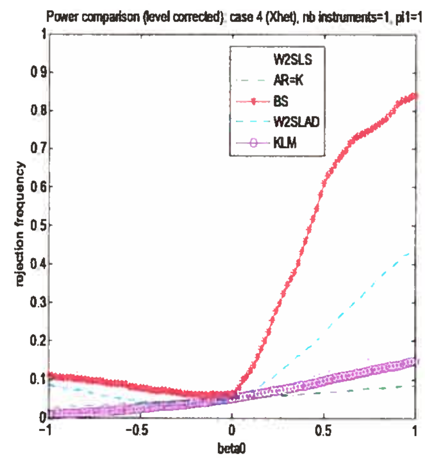
(e) Case 1 weak instrument

(f) Case 2 weak instrument

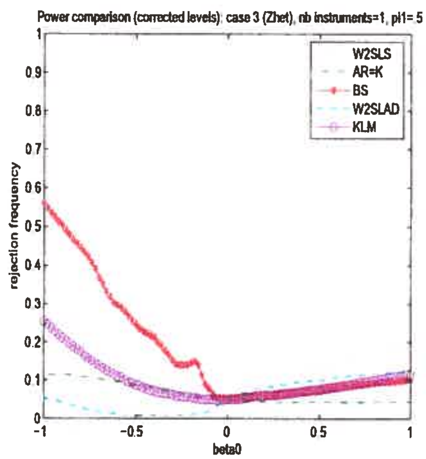
Figure 31. Power functions: model with 5 instruments.



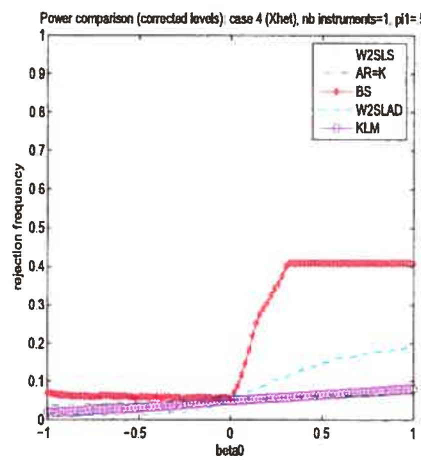
(a) Case 3 strong instrument



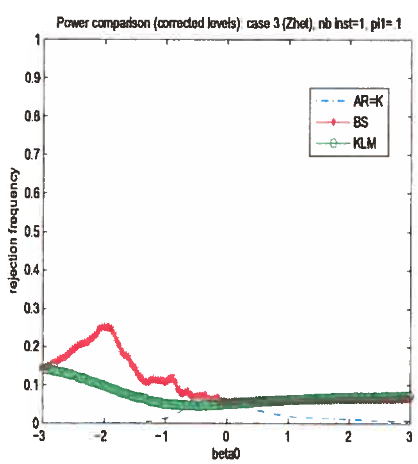
(b) Case 4 strong instrument



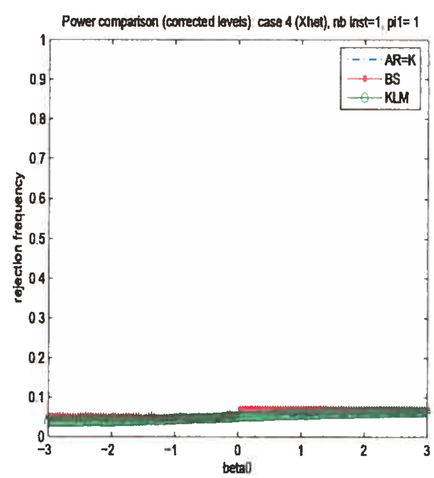
(c) Case 3 instrument .5



(d) Case 4 instrument .5

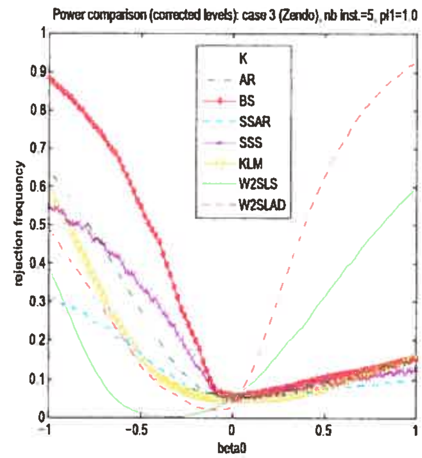


(e) Case 3 weak instrument

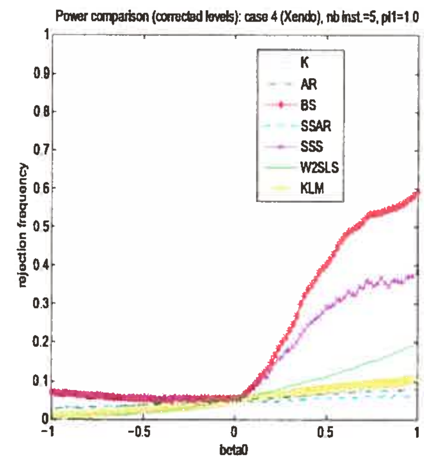


(f) Case 4 weak instrument

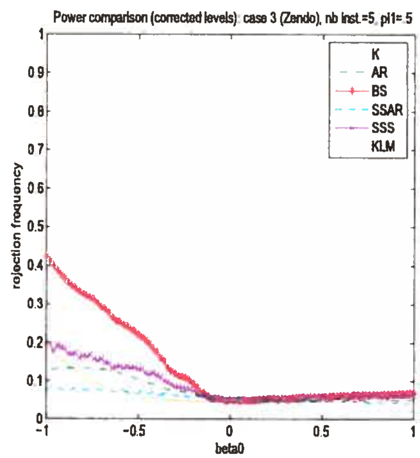
Figure 32. Power functions: model with one instrument: heteroscedastic cases.



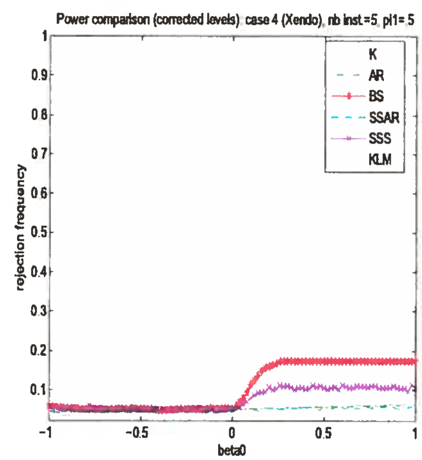
(a) Case 3 strong instrument



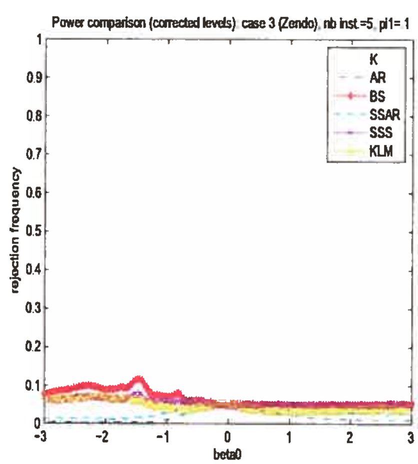
(b) Case 4 strong instrument



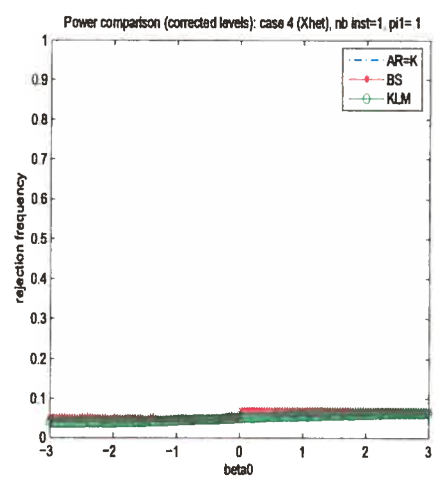
(c) Case 3 instrument .5



(d) Case 4 instrument .5



(e) Case 3 weak instrument



(f) Case 4 weak instrument

Figure 33. Power functions: model with five instruments: heteroscedastic cases.

9. Application: schooling returns

In this section, we apply the method proposed above to study the effect of education on earnings [Angrist and Krueger (1991), Angrist and Krueger (1995), Bound, Jaeger, and Baker (1995), Staiger and Stock (1997), Dufour and Jasiak (2001), Kleibergen (2002, 2005, forthcoming), etc.].[13] Angrist and Krueger (1991) consider an earning equation where the log weekly wage (y_t) is explained by the year number of schooling (x_t) and other covariates (such as the year of birth, age, age squared, race, metropolitan statistical area...). They propose several specifications depending on the included covariates. Further, they use the interactions between the quarter of birth and the year of birth as instruments for correcting the education endogeneity. However, the relation between the instruments and the endogenous variable is apparently weak.

We restrict here on the Angrist and Krueger (1991)'s model specification with dummies for the year of birth as explanatory variables. The data set comes from the 1980 census 5% public-use sample and is composed of $n = 329500$ men born 1930-39.

$$y_i = \beta_0 x_i + \sum_{k=1}^{10} \beta_k d_{ki} + \epsilon_i, \ i = 1, \ldots, n, \tag{9.1}$$

where d_k are dummies for the year of birth. Further, the 30 interactions between the quarter and the year of birth constitute the available "excluded" instruments to correct for the schooling endogeneity. F-statistic for instrument relevance equals 1.573 (with asymptotic p-val= .024), which is low enough to suspect the presence of weak instruments.

We apply split-sample sign-based inference method and compute valid confidence intervals for the education parameter. More precisely, the sample is divided into two parts (1) and (2). With the first part of the sample, we choose the form of quasi-optimal instruments: the year number of schooling is regressed on instruments by OLS. With the second subsample, we construct sign-based statistic using a fitted education. The split-sample sign-based statistics rely on the 11 following moment equations:

$$E[s(y_i^{(2)} - \beta_0 x_i^{(2)} - \sum_{k=1}^{10} \beta_k d_{ki}) \times \tilde{z}_{ji}^{(2)}] = 0, \quad \text{for } i = 1, \ldots, n_2, \ j = 1, \ldots, 11; \tag{9.2}$$

[13] Other questions raised by these data include, for example, the impossibility of a punctual nonparametric identification with discrete instruments [Chesher (2003)] and the problem of many instruments [Hansen, Hausman, and Newey (2005)].

where $\tilde{z}_{ji}^{(2)} = d_{ji},\ j = 1, \ldots, 10$ and $\tilde{z}_{11i}^{(2)}$ is the fitted education. We follow Dufour and Jasiak (2001), and use 10% of the sample for the first stage and 90% for the second one. Two split-sample sign-based statistics are considered. The first one combines moment equations in a classical quadratic GMM form (SSS90). In the second one (TSS90), moment equations are combined following Tippett (1931). Then, Bonferroni-type induced tests are performed using $\alpha_m = \alpha/p$. The idea behind is that quadratic combination of orthogonality conditions refers to some local optimality around the true value of the parameter. In a badly identified setup such as here, other type of combinations like the Tippett's one, may provide better overall properties and smaller confidence intervals.

Table 18 contains 95%$-$confidence intervals obtained with SSS90 and TSS90 but also the Anderson and Rubin statistic (AR), Kleibergen score statistic (K) and Wald (non reliable) CI based on the OLS and the 2SLS estimators. We also report in Table 19 OLS, 2SLS, LIML, SSIV and sign estimates for the return to education.[14]

Projection sign-based confidence intervals obtained using the SSS90 and the TSS90 statistics have smaller spreads than the asymptotic ones based on the AR and K statistics and they are theoretically valid. Moreover, they tend to accept smaller values of the return to education. Table 19 on estimates confirms that point. Sign-based estimates that are very close to 2SLAD estimates, suggest a return to schooling around 4% which is smaller than usually admitted. Such a figure is in adequation with a positive ability bias as expected by the theory.

Then we redo the same experiment on subsamples of 10000 and 2000 observations drawn from the initial sample. We wonder what happens when the sample size gets smaller. Confidence intervals results are reported in Table 20 and estimates in Table 21. We only consider procedures that are robust to weak instruments: K, AR, SSS90 (with 999 replicates) and TSS90 (with 879 replicates).

[14]The CI are smaller than those found by Chernozhukov, Hansen, and Jansson (2006) who exploited a GMM statistic based on the 40 moment equations and included in their model more explanatory variables. We use simulated annealing with different starting points. They used a MCMC algorithm with different starting points.

Table 18. Confidence intervals for schooling returns.

CI	95%	90%	80%
Wald OLS	[.070, .072]	[.071, .072]	[.071, .071]
Wald 2SLS	[.058, .120]	[.063, .115]	[.069, .110]
Wald 2SLAD*	[-.002, .079]	[.004, .073]	[.012, .065]
AR	[.014, .180]	[.022, .169]	[.033, .157]
K	[.054, .133]	[.060, .126]	[.068, .119]
TSS90	[.034, .045]	[.036, .044]	[.037, .043]
SSS90	[.035, .045]	[.036, .041]	[.038, .039]

* W2SLAD CI are obtained by design matrix bootstrap, with 99 replicates [Buchinsky (1998)].

Table 19. Estimates for schooling returns.

	OLS	2SLS	LAD	2SLAD
β_0	.071	.089	.066	.039
	LIML	SSIV90	SSS90	
β_0	.093	.018	.039	

Table 20. Confidence intervals for schooling returns: subsamples n=10000 and n=2000.

CI	95%	90%	80%
		n=10000	
K	[-1,1]	[-1,.222]∪[.239,1]	[-1, -.300]∪[-.012,.145]∪[.404,1]*
AR	[-1,1]	[-.636,.664]	[-.291,.395]
TSS90	[-.190,.109]	[-.110,.083]	[-.034,.049]
SSS90	[-1,1]	[-1,1]	[-1,.236]
		n=2000	
K	[-1,1]	[-1,.073]∪[.106,1]	[-.563, .016]∪[.160,.541]*
AR	[-1,1]	[-1,1]	[-1,.154]∪[.562,1]
TSS90	[-.392,.135]	[-.216,.075]	[-.130,.043]
SSS90	[-1,1]	[-1,1]	[-1,1]

* CIs can be reduced by combining with a J test [Kleibergen (forthcoming)].

Table 21. Estimates for schooling returns: subsamples n=10000 and n=2000.

n=10000

β_0	OLS	2SLS	LAD	2SLAD	LIML	SSIV90	SSS90
	.072	.076	.065	.022	.067	-.012	.022

n=2000

β_0	OLS	2SLS	LAD	2SLAD	LIML	SSIV90	SSS90
	.071	.014	.067	.022	-.119	-.013	.023

The Mincer equation (9.1) sets that the education coefficient has an elasticity form. Consequently, this parameter is constrained in the programs to rely between -1 and 1. Then, a confidence interval of $[-1, 1]$ may refer to an (unconditional) "unbounded" confidence interval. Such a confidence interval indicates a badly identified setup and is in accord with the fact that valid confidence intervals have positive probability to be unbounded in nonidentified setups [Dufour (1997)].

The CI spread based on SSS90 and AR statistics increases as the number of observations decreases. 90%-CI based on the AR statistic is bounded for $n = 10000$ whereas for $n = 2000$, the 90%-CI is $[-1, 1]$. The same occurs with 95%-CI based on the SSS90 statistic. The behavior of the K statistic is less clear. As it is a quadratic form of the score of the concentrated log-likelihood, it basically contains information on a slope. Its use is locally justified around the LIML estimator but may follow a somewhat odd behavior outside that neighborhood. The Tippett-sign-based statistic provides the smaller CIs for both subsamples, which indicates that quadratic combinations of orthogonality conditions are not optimal in small subsamples.

Concerning estimates (Table 21), our findings are similar to the whole sample ones. Sign-based estimates are very close to 2SLAD estimates and suggest returns to schooling around 2% in both subsamples which is in adequation with the theoretically expected ability bias.

10. Conclusion

In this paper, we presented a finite-sample sign-based inference system for the parameter of a structural possibly nonlinear model. We introduced a condition of instrument validity with respect to the signs of the structural error. We showed that, under the instruments validity, the distribution of the structural error sign vector is known and does not depend on any nuisance parameter. This allowed us to conduct a Monte Carlo-based inference using on the exact distribution of IV sign-based statistics. The derived joint tests are exact for any sample size and are robust to identification failures. Tests of more general hypothesis and confidence sets are then constructed using projection techniques. Our approach is in the spirit of Anderson and Rubin (1949). The IV sign-based statistics we studied can be constructed from auxiliary regressions of the constrained signs on auxiliary instruments. We also considered the problem of approaching the optimal set of instruments to include in the model in case of overidentification using two different optimality concepts (point and local optimality). Finally, IV sign-based estimators are presented. They turn to be consistent and asymptotically normal when identification holds under weaker assumptions than the ones required in the 2SLAD asymptotic theory. Besides, they can directly be associated with previous sign-based inference, which avoids one to use complicated methods such as the bootstrap. By construction, the level of IV sign-based tests is controlled and simulations indicate that those tests perform better than usual ones (including methods that are robust to weak instruments or identification failures) in finite samples, when the data are heterogenous, heteroskedastic or when endogenous variables affect the structural error distribution without affecting its sign. Finally, sign-based inference is applied to the Angrist and Krueger's returns to schooling problem. Sign-based estimate of the return to schooling is around 4% and projection-based confidence intervals, besides being more robust, are more precise than those based on the AR or the K statistics. In small samples, it seems that Tippett-type combination of orthogonality conditions provides better properties than usual quadratic combination and leads to more precise confidence intervals.

Appendix

A. Proofs

A.1. Proof of Proposition 3.1

Consider the vector $[s(u_1), s(u_2), \ldots, s(u_n)]' \equiv (s_1, s_2, \ldots, s_n)'$. From Assumption A1, we derive the two following equalities:

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

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

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

Hence, conditional on Z, $s_1, s_2, \ldots, s_n$ are distributed like n i.i.d random variables with distribution:

$$P(s_t = 1) = P(s_t = -1) = \frac{1}{2},\ t = 1, \ldots, n.$$

A.2. Proof of Proposition 4.1

This is a direct application of Neyman-Pearson lemma. The likelihood function of S under H_0 is

$$L_0(s_1, \ldots, s_n) = \prod_{t=1}^{n} p_{0t}^{(1+s_t)/2} p_{0t}^{(1-s_t)/2}$$

and under H_1,

$$L_1(s_1, \ldots, s_n) = \prod_{t=1}^{n} p_{1t}^{(1+s_t)/2} p_{1t}^{(1-s_t)/2}.$$

Hence, after some computations, the loglikelihood ratio becomes

$$\ln\left(\frac{L_1}{L_0}\right) = \sum_{t=1}^{n}(1/2)\left[\ln\left(\frac{p_{1t}(1-p_{1t})}{p_{0t}(1-p_{0t})}\right) + s_t \ln\left(\frac{p_{1t}(1-p_{0t})}{p_{0t}(1-p_{1t})}\right)\right], \quad \text{(A.1)}$$

and yields the optimal test against H_1. The most powerful test based on S rejects H_0 when

$$\sum_{t=1}^{n} s_t \ln\left(\frac{p_{1t}(1-p_{0t})}{p_{0t}(1-p_{1t})}\right) > c(\alpha, H_1)$$

where $c(\alpha, H_1) = c - \sum_{t=1}^{T}(1/2)\left[\ln\left(\frac{p_{1t}(1-p_{1t})}{p_{0t}(1-p_{0t})}\right)\right]$ with c derived from Neyman-Pearson condition. □

A.3. Proof of Corollary 4.2

In the regression framework, $(p_{01}, \ldots, p_{0n})$ and $(p_{11}, \ldots, p_{1n})$ are known. As Assumption A1 holds under H_0, we have $p_{0t} = .5$, and under H_1, we can write for $t = 1, \ldots, n.$,

$$p_{1t} = P_{H_1}[f(y_t, x_t, \theta_0) > 0] = P_{H_1}[f(y_t, x_t, \theta_1) > f(y_t, x_t, \theta_1) - f(y_t, x_t, \theta_0)] = 1 - G(h_t),$$

where $h_t = f(y_t, x_t, \theta_0) - f(y_t, x_t, \theta_1)$. Hence, the point-optimal sign-based test of H_0 against H_1 rejects H_0 when

$$\sum_{t=1}^{n} s(u_t) \ln\left(\frac{1 - G(h_t)}{G(h_t)}\right) > c(\alpha, \theta_1), \quad \text{(A.2)}$$

where $(h_1, \ldots, h_n)' = \big(f(y_1, x_1, \theta_1) - f(y_1, x_1, \theta_0), \ldots, f(y_n, x_n, \theta_1) - f(y_n, x_n, \theta_0)\big)'$ and $c(\alpha, \theta_1)$ depending on the level.

A.4. Proof of Proposition 5.1

First, we prove the following lemma.

Lemma A.1 *Let $\{G_n\}_n$ be a sequence of real functions tending uniformly towards G on a compact set $K \subset \mathbb{R}$ and $0 \in int(K)$. Suppose further that G_n and G are differentiable with continuous derivative on K for all n and satisfy $n(G_n(0) - G(0)) \to 0$ and $G_n'(0) - G'(0) \to 0$. Then,*

$$\sup_{y \in B(0, \frac{1}{n})} ||G_n(y) - G(y)|| = o(1/n).$$

Proof of Lemma A.1. Taylor expansions gives

$$G_n(x) = G_n(0) + xG_n'(0) + o(|x|), \forall x \in B(0, 1/n) \cap K, \tag{A.3}$$

and

$$G(x) = G(0) + xG'(0) + o(|x|), \forall x \in B(0, 1/n) \cap K. \tag{A.4}$$

We can write

$$|G_n(x) - G(x)| = |G_n(0) - G(0) + x(G_n'(0) - G'(0)) + o(1/n)|. \tag{A.5}$$

Hence,

$$|G_n(x) - G(x)| \leq |G_n(0) - G(0)| + \frac{1}{n}|G_n'(0) - G'(0)| + o(1/n) \tag{A.6}$$

by majoring $|x|$ by $1/n$. That entails

$$|G_n(x) - G(x)| = o(1/n).\square \tag{A.7}$$

Let us now consider the problem of testing $H_0 : \theta = \theta_0$ against alternatives of the general form $H_1 : \theta = \theta_1$. The power function of a sign-based test T conditional on Z is

$$\beta(\theta_1) = P_{\theta_1}\left[T\big(s(f(y, x, \theta_0)), Z\big) > 1 - c_T(Z, \alpha)|Z\right] = P_{\theta_1}[S \in W_\alpha|Z] \tag{A.8}$$

where S is the random variable of the constrained signs and W_α the critical region of the test with level α. In the sequel, we omit to write that all results are conditional on Z. To identify the instruments which maximize the power function in the neighborhood of θ_0, we first derive the sign distribution under H_1. The independence assumption implies that the sign distribution is the product of terms of the form

$$P_{\theta_1}[s_t = s] = P_{\theta_1}[f(y_t, x_t, \theta_0) \geq 0]^{\frac{1+s}{2}} P_{\theta_1}[f(y_t, x_t, \theta_0) < 0]^{\frac{1-s}{2}}. \tag{A.9}$$

As f is continuously differentiable, the mean value theorem entails

$$f(y_t, x_t, \theta_1) = f(y_t, x_t, \theta_0) + \left.\frac{\partial f(y_t, x_t, \theta)}{\partial \theta'}\right|_{\theta=\bar{\theta}_t} (\theta_1 - \theta_0),\ t = 1, \ldots, n, \tag{A.10}$$

where $\bar{\theta}_t = p_t\theta_0 + (1 - p_t)\theta_1$ with $p_t = p_t(y_t, x_t, \theta_0, \theta_1) \in [0, 1],\ t = 1, \ldots, n.$ Let us denote

$$H_t(\bar{\theta}_t) = \left.\frac{\partial f(y_t, x_t, \theta)}{\partial \theta'}\right|_{\theta=\bar{\theta}_t},\ t = 1, \ldots, n. \tag{A.11}$$

We can rewrite

$$f(y_t, x_t, \theta_0) = f(y_t, x_t, \theta_1) - [H_t(\overline{\theta}_t) - EH_t(\overline{\theta}_t)](\theta_1 - \theta_0) - EH_t(\overline{\theta}_t)(\theta_1 - \theta_0). \quad (A.12)$$

This yields, using equation (A.9)

$$\begin{aligned} P_{\theta_1}[s_t = s] &= P_{\theta_1}\left[u_t - \left(H_t(\overline{\theta}_t) - EH_t(\overline{\theta}_t)\right)(\theta_1 - \theta_0) > EH_t(\overline{\theta}_t)(\theta_1 - \theta_0)\right]^{\frac{1+s}{2}} \\ &\times \; P_{\theta_1}\left[u_t - \left(H_t(\overline{\theta}_t) - EH_t(\overline{\theta}_t)\right)(\theta_1 - \theta_0) \leq EH_t(\overline{\theta}_t)(\theta_1 - \theta_0)\right]^{\frac{1-s}{2}}. \end{aligned}$$

As the observations are $i.i.d.$, we will not write the subscript t. Let us denote

$$G_n^{\overline{\theta}_n}(x) = P_{\theta_n}\left[u - \left(H(\overline{\theta}_n) - EH(\overline{\theta}_n)\right)(\theta_n - \theta_0) \leq x\right] \quad (A.13)$$

where the *real* random variable $u \sim G$. Equation (A.13) can alternatively be written

$$P_{\theta_n}[s = s_a] = \left[\frac{1}{2} - G_n^{\overline{\theta}_n}\left(EH(\overline{\theta}_n)(\theta_n - \theta_0)\right)\right] s_a + \frac{1}{2} \quad (A.14)$$

where again s stands for a *real* random variable and not for a vector.

Let us now examine

$$\begin{aligned} R &= G_n^{\overline{\theta}_n}\left(EH(\overline{\theta}_n)'(\theta_n - \theta_0)\right) - G\left(EH(\overline{\theta}_n)'(\theta_n - \theta_0)\right) && (A.15) \\ &+ G\left(EH(\overline{\theta}_n)'(\theta_n - \theta_0)\right) - G(0) - G'(0)EH(\overline{\theta}_n)'(\theta_n - \theta_0) \\ &- \frac{1}{2}G''(0)(\theta_n - \theta_0)'EH(\overline{\theta}_n)EH(\overline{\theta}_n)'(\theta_n - \theta_0). && (A.16) \end{aligned}$$

When $\theta_n \longrightarrow \theta_0$, we want to show that R is $o(||\theta_0 - \theta_n||^2)$. For this, we denote:

$$A = G_n^{\overline{\theta}_n}\left(EH(\overline{\theta}_n)'(\theta_n - \theta_0)\right) - G\left(EH(\overline{\theta}_n)'(\theta_n - \theta_0)\right),$$

$$\begin{aligned} B &= G\left(EH(\overline{\theta}_n)'(\theta_n - \theta_0)\right) - G(0) - G'(0)EH(\overline{\theta}_n)'(\theta_n - \theta_0) \\ &- \frac{1}{2}G''(0)(\theta_n - \theta_0)'EH(\overline{\theta}_n)EH(\overline{\theta}_n)'(\theta_n - \theta_0). \end{aligned}$$

We first consider B. We easily have

$$||B|| = o(||\theta_n - \theta_0||^2) \quad (A.17)$$

using a Taylor expansion of G in the vicinity of zero, because $EH(\overline{\theta}_n)$ is uniformly bounded by M_1 around θ_0 (condition **c**). Let us consider now A. We can major $||A||$ by

$$||A|| \leq M_1||\theta_n - \theta_0|| \sup_{y \in B(0, M_1||\theta_0 - \theta_n||)} ||G(y) - G_n(y)||. \quad (A.18)$$

Moreover, as $\{G_n\}_{n\in\mathbb{N}}$ are increasing continuous functions that converge everywhere to G, a Dini-type theorem implies the convergence is uniform. Hence, Lemma **A.1** applies. Finally

$$\sup_{y\in B(0,M_1||\theta_0-\theta_n||)} ||G(y)-G_n(y)|| = o(||\theta_n-\theta_0||). \tag{A.19}$$

Finally

$$||A|| = o(||\theta_n-\theta_0||^2). \tag{A.20}$$

Consequently, inequalities (A.20) and (A.17) with condition **d** entail:

$$P_{\theta_n}[s_t=s]-\frac{1}{2} = s\big[-G'(0)\big(EH_t(\theta_n)'(\theta_n-\theta_0)\big)+o(||\theta_n-\theta_0||^2)\big]. \tag{A.21}$$

As $(s_1,\ldots,s_n)$ are $i.i.d.$, it follows

$$\begin{aligned} P_{\theta_n}[S=(s_1,\ldots,s_n)] &= \left(\frac{1}{2}\right)^n-\left(\frac{1}{2}\right)^{n-1}\sum_{t=1}^{n}s_t[G'(0)\big(EH_t(\theta_n)'(\theta_n-\theta_0)\big)] \\ &- \left(\frac{1}{2}\right)^{n-2}\sum_{t\le l}s_ts_l[G'(0)^2(\theta_n-\theta_0)(EH_t(\theta_n)')(EH_l(\theta_n))(\theta_n-\theta_0')] \\ &+ o(||\theta_n-\theta_0||^2)]. \end{aligned} \tag{A.22}$$

The remainder follows the proof of Proposition 4.1 in Coudin and Dufour (2005a) and Boldin, Simonova, and Tyurin (1997). We consider sign-based tests that maximize the mean curvature around θ_0. It is trivial to see that the locally optimal test with critical region W_α is locally unbiased (assuming the opposite goes to a contradiction), $i.e.$

$$\left.\frac{d\mathrm{P}_\theta[W_\alpha]}{d\theta}\right|_{\theta=\theta_0} = 0. \tag{A.23}$$

The behavior of the power function around zero is then totally defined by the quadratic term of its Taylor expansion which can be identified thanks to equation (A.22). The mean curvature is by definition proportional to the trace of $\frac{d^2\mathrm{P}_\theta[W]}{d\theta^2}$ at $\theta=\theta_0$ [see Boldin, Simonova, and Tyurin (1997), p. 41, Dubrovin, Fomenko, and Novikov (1984), Chapter 2, pp. 76-86 or Gray (1998), Chapter 21, pp. 373-380]. Taking the trace in the expression of equation (A.22), we find (after some computations) it is proportional to

$$\sum_{1\le t\neq}\sum_{l\le n}G'(0)^2s_ts_lEH_t(\theta_0)EH_l(\theta_0)'. \tag{A.24}$$

By adding the quantity $\sum_{t=1}^{n}\left(EH_t(\theta_n)EH_t(\theta_n)'\right)$ to (A.24), we find the locally optimal sign-based test in the sense proposed by Boldin, Simonova, and Tyurin (1997) is

$$W = \left\{s : s'(y)\left[EH(\theta_0)EH(\theta_0)\right]'s(y) > c'_\alpha\right\} . \tag{A.25}$$

Standardizing by $EH(\theta_0)'EH(\theta_0)$ then leads to

$$W = \{s : s'(y)EH(\theta_0)[EH(\theta_0)'EH(\theta_0)]^{-1}EH(\theta_0)'s(y) > c'_\alpha\} . \tag{A.26}$$

A.5. Proof of Theorem 7.1 (Consistency)

Consistency of IV sign-estimators is an extension of consistency of classical sign estimators [Theorem 5.9 in Coudin and Dufour (2005b)]. Both proofs follow the same classical 4 steps (pointwise convergence, weak uniform convergence, consistency and identification). Here, we indicate only points that differ. The stochastic process considered here is $W^v = \{W_t^v = (y_t,\ x_t',\ v_t')\}_{t=1,2,\dots} : \Omega \to \mathbb{R}^{p+k+l}$, and we denote

$$q_t(w_t, \theta) = s\big(f(y_t, x_t, \theta)\big) \otimes h_t(v_t, \theta),\ t = 1, \dots, n, \tag{A.27}$$

which satisfies the same mixing condition. Similarly to Theorem 5.9 in Coudin and Dufour (2005b), pointwise convergence for any θ is implied by assumptions A7, A9 (boundedness point) and Corollary 3.48 of White (2001).

Uniform convergence and continuity of the limiting function are implied by the generic law of large number of Andrews (1987). Andrew's conditions B1, B2 and A1 are fulfilled by assumptions A7, A8, A9 and A10. Furthermore, we use his comment 3 to conclude on the weak continuity condition (A6). Condition A6(a) allows $q_t(w,\theta)$ to have isolated discontinuities provided $q_t(w,\theta)p_t(w)$ is continuous in θ uniformly in t a.e.$[\mu]$, where μ is a σ-finite measure, that dominates each of the marginal distribution of $W_t,\ t = 1, 2\dots$ and $p_t(w)$ is the density of W_t w.r.t. μ. Condition A6(b) states that $\int \sup_{t\geq 1}|q_t(w,\theta)|p_t(w)d\mu(w) < \infty$. Here, we consider $\mu = P$, $q_t(w,\theta)p_t(w)$ is continuous in θ a.e. w.r.t. P, as p_t does not depend on θ and q_t is a continuous function everywhere except at $\{f(y_t, x_t, \theta) = 0\}$ which is a P-negligible set: $P[\{w\ :\ f(y_t, x_t, \theta) = 0\}] = 0$ (no tie assumption A8). Furthermore, q_t is L_1-bounded and uniformly integrable. Then, condition A6 is fulfilled. The consistency part applies without further modifications. Finally, the identification conditions A11 and A12 allow to conclude on consistency.

A.6. Proof of Theorem 7.2 (Asymptotic Normality)

If $z_t = h_t(\theta, v_t)$, Assumptions A9, A17, A14 and A15 allow to differentiate below the integral.

$$\begin{aligned}\frac{\partial}{\partial\theta'}E[h_t(\theta,v_t)s(f(y_t,x_t,\theta))] &= E\left[h_t(\theta,v_t)\frac{\partial f(y_t,x_t,\theta)}{\partial\theta'}g_t\big(f(y_t,x_t,\theta)|z_1,\ldots,z_n\big)\right]\\ &\quad +E\left[\frac{\partial h_t(\theta,v_t)}{\partial\theta'}\right]s\left(f(y_t,x_t,\theta)\right). \qquad \text{(A.28)}\end{aligned}$$

By uniform convergence (shown in the consistency part), it follows that the limiting objective function, $\lim_n \frac{1}{n}\sum_{t=0}^{n} E\big[z_t(\theta)s\big(f(y_t,x_t,\theta)\big)\big]$, is differentiable with derivative $L(\theta)$:

$$\begin{aligned}L(\theta) &= \lim_{n\to\infty}\frac{1}{n}\sum_t E\left[h_t(\theta,v_t)\frac{\partial f(y_t,x_t,\theta)}{\partial\theta'}g_t\big(f(y_t,x_t,\theta)|z_1,\ldots,z_n\big)\right]\\ &\quad +\frac{1}{n}\sum_t E\left[\frac{\partial h_t}{\partial\theta'}s\big(f(y_t,x_t,\theta)\big)\right].\end{aligned}$$

Theorem 7.2 in Newey and McFadden (1994) may then be applied. Their condition (i), which states that 0 is attained at the limit by θ_0, is fulfilled by the moment condition A3. Their condition (ii) states that the limit objective function is differentiable at θ_0 and positive definite. This is fulfilled by the first part of our proof and condition A19. Then, their condition (iii) (interior) is implied by A10. Using the mixing specification A16 of $\{w\}$ and conditions A3, A9, A13 and A18, we apply a White-Domowitz central limit theorem [see White (2001), Theorem 5.20]. This fulfills condition (iv) of Theorem 7.2 in Newey and McFadden (1994). Finally, condition v (stochastic equicontinuity) is implied by uniform convergence (see the consistency part) which completes the proof.

Références

ABDELKHALEK, T., AND J.-M. DUFOUR (1998) : “Statistical Inference for Computable General Equilibrium Models, with Application to a Model of the Moroccan Economy,” *Review of Economics and Statistics*, 80(4), 520–534.

AMEMIYA, T. (1982) : “Two-Stage Least Absolute Deviations Estimator,” *Econometrica*, 50(3), 689–711.

ANDERSON, T. W., AND H. RUBIN (1949) : “Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations,” *Annals of Mathematical Statistics*, 20(1), 46–63.

ANDREWS, D. (1991) : “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,” *Econometrica*, 59(3), 817–858.

ANDREWS, D., AND V. MARMER (2005) : “Exactly Distribution-Free Inference in Instrumental Variables Regression with Possible Weak Instruments,” Discussion paper CFDP 1501, Cowles Foundation for Research in Economics.

ANDREWS, D., M. MOREIRA, AND J. STOCK (2004) : “Optimal Invariant Similar Tests for Instrumental Variables Regression,” Discussion paper T0299, N.B.E.R.

——— (2006) : “Optimal Two-sided Invariant Similar Tests for Instrumental Variables Regression,” *Econometrica*, 3(70), 715–752.

ANDREWS, D., AND J. STOCK (2005) : “Inference with Weak Instruments,” Discussion paper, Econometric Society World Meeting 2005, Invited Session.

ANDREWS, D. W. (1987) : “Consistency in Nonlinear Econometric Models : A Generic Uniform Law of Large Numbers,” *Econometrica*, 55(6), 1465–1471.

ANGRIST, J., AND A. KRUEGER (1991) : “Does Compulsory School Attendance Affect Schooling and Earning ?,” *The Quarterly Journal of Economics*, 106(4), 979–1014.

——— (1995) : “Split-Sample Instrumental Variables Estimates of the Return to Schooling,” *Journal of Business and Economic Statistics*, 13(2), 225–235.

BAHADUR, R., AND L. SAVAGE (1956) : "The Nonexistence of Certain Statistical Procedures in Nonparametric Problem," *Annals of Mathematical Statistics*, 27(4), 1115–1122.

BARNARD, G. A. (1963) : "Comment on 'The Spectral Analysis of Point Processes' by M. S. Bartlett," *Journal of the Royal Statistical Society, Series B*, 25(2), 294.

BARRO, R., AND X. SALA-I MARTIN (1991) : "Convergence Across States and Regions," *Brookings Papers on Economic Activity*, 1991(1), 107–182.

BEKKER, P., AND F. KLEIBERGEN (2001) : "Finite-Sample Instrumental Variables Inference Using an Asymptotically Pivotal Statistic," Discussion paper 01-055/4, Tinbergen Institute.

BEKKER, P., AND S. LAWFORD (2005) : "Symmetry-Based Inference in an Instrumental Setting," Discussion paper, Department of Economics, University of Groningen.

BOLDIN, M. V., G. I. SIMONOVA, AND Y. N. TYURIN (1997) : *Sign-Based Methods in Linear Statistical Models*, vol. 162 of *Translations of Mathematical Monographs*. American Mathematical Society, Maryland.

BOUND, J., D. JAEGER, AND R. BAKER (1995) : "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variables is Weak," *Journal of the American Statistical Association*, 90(430), 443–450.

BUCHINSKY, M. (1994) : "Changes in the U.S. Wages Stucture 1963-1987 : Exploration by Quantiles Regressions," *Econometrica*, 62(2), 405–458.

——— (1995) : "Estimating the Asymptotic Covariance Matrix for Quantile Regression Models," *Journal of Econometrics*, 68(2), 303–338.

——— (1998) : "Recent Advances in Quantile Regression Models : A Practical Guideline for Empirical Research," *Journal of Human Resources*, 33(1), 88–126.

BUCHINSKY, M., AND H. J. (1998) : "An Alternative Estimator for the Censored Quantile Regression Model," *Econometrica*, 66(3), 653–671.

CAMPBELL, B., AND J.-M. DUFOUR (1991) : "Over-rejections in Rational Expectations Models : A Nonparametric Approach to the Mankiw-Shapiro Problem," *Economics Letters*, 35(3), 285–290.

——— (1995) : "Exact Nonparametric Orthogonality and Random Walk Tests," *Review of Economics and Statistics*, 77(1), 1–16.

——— (1997) : "Exact Nonparametric Tests of Orthogonality and Random Walk in the Presence of a Drift Parameter," *International Economic Review*, 38(1), 151–173.

CHAMBERLAIN, G. (1994) : "Quantile Regression, Censoring and the Structure of Wages," in *Advances in Econometrics*, ed. by C. Sims, pp. 171–209, New-York. Elsevier.

CHEN, L., AND S. PORTNOY (1996) : "Two-Stage Regression Quantiles and Two-Stage Trimmed Least Squares Estimators for Structural Equation Models," *Communications in Statistics- Theory and Methods*, 25, 1005–1032.

CHEN, X., O. LINTON, AND I. VAN KEILEGOM (2003) : "Estimation of Semiparametric Models When the Criterion Function is Not Smooth," *Econometrica*, 71(5), 1591–1608.

CHERNOZHUKOV, V., AND C. HANSEN (2004) : "Instrumental Variable Quantile Regression," Discussion paper 06-19, M.I.T. Department of Economics.

CHERNOZHUKOV, V., C. HANSEN, AND M. JANSSON (2006) : "Finite Sample Inference for Quantile Regression Models," Discussion paper 06-03, M.I.T. Department of Economics.

CHERNOZHUKOV, V., E. TAMER, AND H. HONG (2006) : "Inference on Parameter Sets in Econometric Models," Discussion paper 06-18, M.I.T. Department of Economics.

CHESHER, A. (2003) : "Nonparametric Identification under Discrete Variation," Discussion paper CWP 19/03, Cemmap, Institute for Fiscal Studies and Department of Economics, University College of London.

CHOW, Y., AND H. TEICHER (1988) : *Probability Theory : Independence, Interchangeability, Martingales*, Springer Texts in Statistics. Springer-Verlag, New-york, 2nd edn.

CHUNG, K. (2001) : *A Course in Probability Theory*. Academic Press, New-York.

COAKLEY, C. W., AND M. A. HEISE (1996) : "Versions of the Sign Test in the Presence of Ties," *Biometrics*, 52(4), 1242–1251.

COUDIN, E., AND J.-M. DUFOUR (2005a) : "Finite and Large Sample Distribution-Free Inference in Linear Median Regressions under Heteroskedasticity and Nonlinear Dependence of Unknown Form," Discussion paper, CIREQ, Université de Montréal and CREST-INSEE.

——— (2005b) : "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.

CRUZ, L., AND M. MOREIRA (2005) : "On the Validity of Econometric Techniques with Weak Instruments," *The Journal Of Human Resources*, 40(2), 393–410.

DAVIDSON, R., AND J. MCKINNON (2001) : "Artificial regressions," Discussion paper 1038, Queen's University, Department of Economics.

DE ANGELIS, D., P. HALL, AND G. A. YOUNG (1993) : "Analytical and Bootstrap Approximations to Estimator Distributions in $L1$ Regression," *Journal of the American Statistical Association*, 88(424), 1310–1316.

DIELMAN, T., AND R. PFAFFENBERGER (1988a) : "Bootstrapping in Least Absolute Value Regression : An Application to Hypothesis Testing," *Communications in Statistics-Simulation and Computation*, 17(3), 843–856.

——— (1988b) : "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*, no. 31 in Lecture Notes - Monograph Series. Institute of Mathematical Statistics, Hayward, CA.

DUBROVIN, B., A. FOMENKO, AND S. NOVIKOV (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(3), 117–128.

——— (1990) : "Exact Tests and Confidence Sets in Linear Regressions with Autocorrelated Errors," *Econometrica*, 58(2), 475–494.

——— (1997) : "Some Impossibility Theorems in Econometrics, with Applications to Structural and Dynamic Models," *Econometrica*, 65(6), 1365–1389.

——— (2003) : "Identification, Weak Instruments and Statistical Inference in Econometrics. Presidential Address to the Canadian Economics Association," *Canadian Journal of Economics*, 36(4), 767–808.

——— (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 J. JASIAK (2001) : "Finite Sample Limited Information Inference Methods for Structural Equations and Models with Generated Regressors," *International Economic Review*, 42(3), 815–843.

DUFOUR, J.-M., AND J. KIVIET (1998) : "Exact Inference Methods for First-order Autoregressive Distributed Lag Models," *Econometrica*, 82(1), 79–104.

DUFOUR, J.-M., AND A. TAAMOUTI (2006) : "Exact Optimal and Adaptative Inference in Linear and Nonlinear Models Under Heteroskedasticity and Non-Normality of Unknown Forms," Discussion paper, Université de Montréal.

DUFOUR, J.-M., AND M. TAAMOUTI (2002) : "Point-Optimal Instruments and Generalized Anderson-Rubin procedures for Nonlinear models," Discussion paper, C.R.D.E., Université de Montréal.

——— (2005) : "Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments," *Econometrica*, 73(4), 1351–1365.

DWASS, M. (1957) : "Modified Randomization Tests for Nonparametric Hypotheses," *Annals of Mathematical Statistics*, 28(1), 181–187.

EFRON, B. (1998) : "R. A. Fisher in the 21st Century," *Statistical Science*, 13(2), 95–122.

ENGLE, R., AND S. MANGANELLI (1999) : "CAViaR : Conditional Autoregressive Value at Risk by Regression Quantiles," Discussion paper 99-20, Department of Economics, UC San Diego.

ENGLE, R. F., AND D. L. MCFADDEN (eds.) (1994) : *Handbook of Econometrics*, vol. 4. Elsevier, North Holland.

FISHER, R. A. (1930) : "Inverse Probability," *Proc. Cambrige Philos. Soc.*, 26, 528–535.

FITZENBERGER, B. (1997a) : "A Guide to Censored Quantile Regressions," in *Handbook of Statistics*, ed. by G. S. Maddala, and C. R. Rao, vol. 15, chap. 6, pp. 405–437. North-Holland, North Holland.

——— (1997b) : "The Moving Blocks Bootstrap and Robust Inference for Linear Least Squares and Quantile Regressions," *Journal of Econometrics*, 82(2), 235–287.

GALLANT, A., D. HSIEH, AND G. TAUCHEN (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., G. D. FERRIER, AND J. ROGERS (1994) : "Global Optimization of Statistical Functions with Simulated Annealing," *Journal of Econometrics*, 60(1), 65–100.

GOURIÉROUX, C., AND A. MONFORT (1995a) : *Statistics and Econometric Models*, vol. 1. Cambridge University Press.

——— (1995b) : *Statistics and Econometric Models*, vol. 2. Cambridge University Press.

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

HAHN, J. (1997) : "Bayesian Boostrap of the Quantile Regression Estimator : A Large Sample Study," *International Economic Review*, 38(4), 795–808.

HANNIG, J. (2006) : "On Fiducial Inference – the good, the bad and the ugly," Discussion paper, Department of Statistics, Colorado State University.

HANSEN, C., J. HAUSMAN, AND W. NEWEY (2005) : "Estimation with Many Instrumental Variables," Discussion paper, M.I.T., Department of Economics.

HANSEN, L. (1982) : "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50(4), 1029–1054.

HAUSMAN, J. (1983) : "Specification and Estimaton of Simultaneous Models," in *Handbook of Econometrics*, ed. by Z. Grilichest, and M. Intriligator, vol. 1, chap. 7, pp. 391–448. Marcel Dekker, New York.

HODGES, J. L., AND E. L. LEHMANN (1963) : "Estimates of Location Based on Rank Tests," *The Annals of Mathematical Statistics*, 34(2), 598–611.

HONG, H., AND E. TAMER (2003) : "Inference in Censored Models With Endogenous Regressors," *Econometrica*, 71(3), 905–932.

HONORE, B., AND L. HU (2004) : "On the Performance of Some Robust Instrumental Variables," *Journal of Business & Economic Statistics*, 22(1), 30–39.

HOROWITZ, J. (1998) : "Bootstrap Methods for Median Regression Models," *Econometrica*, 66(6), 1327–1352.

JAECKEL, L. (1972) : "Estimating Regression Coefficients by Minimizing the Dispersion of the Residuals," *The Annals of Mathematical Statistics*, 40(5), 1449–1458.

JOHNSON, N., S. KOTZ, AND C. READ (eds.) (1983) : *Encyclopedia of Statistical Sciences*, vol. 3. John Wiley & Sons, New York.

JUNG, S. (1996) : "Quasi-Likelihood for Median Regression Models," *Journal of the American Statistical Association*, 91(433), 251–257.

JURECKOVA, J. (1971) : "NonParametric Estimate of Regression Coefficients.," *The Annals of Mathematical Statistics*, 42(4), 1328–1338.

KIM, T., AND H. WHITE (2002) : "Estimation, Inference, and Specification Testing for Possibly Misspecified Quantile Regression," Discussion paper 2002-09, Department of Economics, UC San Diego.

KLEIBERGEN, F. (2002) : "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70(5), 1781–1803.

——— (2005) : "Testing Parameters in GMM Without Assuming They Are Identified," *Econometrica*, 73(4), 1103–1123.

——— (forthcoming) : "Generalizing Weak Instrument Robust IV Statistics Towards Multiple Parameters, Unrestricted Covariance Matrices and Identification Statistics," *Journal of Econometrics*.

KOENKER, R., AND G. BASSETT (1982) : "Tests of Linear Hypotheses and L_1 Estimation," *Econometrica*, 50(6), 1577–1584.

KOENKER, R., AND G. BASSETT, JR. (1978) : "Regression Quantiles," *Econometrica*, 46(1), 33–50.

KOENKER, R., AND K. HALLOCK (2000) : "Quantile Regression : An Introduction," Discussion paper, University of Illinois.

——— (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.

KOUL, H. (1971) : "Asymptotic Behavior of a Class of Confidence Regions Based on Ranks in Regression.," *The Annals of Mathematical Statistics*, 42(2), 466–476.

LEE, S. (2003) : "Efficient Semiparametric Estimation of a Partially Linear Quantile Regression Model," *Econometric Theory*, 19(1), 1–31.

LEHMANN, E. (1959) : *Testing Statistical Hypothesis*. John Wiley, New York, first edition.

LEHMANN, E. L., AND C. STEIN (1949) : "On the Theory of some Non-Parametric Hypotheses," *Annals of Mathematical Statistics*, 20(1), 28–45.

MANKIW, G., AND M. SHAPIRO (1986) : "Do We Reject too Often ? Small Samples properties of Tests of Rational Expectations Models," *Economics Letters*, 20(2), 139–145.

MOREIRA, M. (2001) : "Tests with Correct Size When Instruments can be Arbitrary Weak," Discussion paper 37, Center for Labor Economics, UC Berkeley.

——— (2003) : "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71(4), 1027–1048.

MOREIRA, M., AND B. POI (2003) : "Implementing Conditional Tests with Correct Size in the Simultaneous Equations Model," *The Stata Journal*, 3(1), 57–70.

NELSON, C., R. STARTZ, AND E. ZIVOT (1998) : "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," *International Economic Review*, 39(4), 1119–1144, Symposium on Forcasting and Empirical Methods in Macroeconomics and Finance.

NEWEY, W., AND K. WEST (1987) : "A Simple Positive Semi-Definite, Heteroskadasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.

NEWEY, W. K., AND D. MCFADDEN (1994) : "Large Sample Estimation and Hypothesis Testing," in Engle and McFadden (1994), chap. 36, pp. 2113–2245.

NEYMAN, J. (1941) : "Fiducial argument and the theory of confidence intervals," *Biometrika*, 32(2), 128–150.

PARZEN, E. (1957) : "On Consistent Estimates of the Spectrum of a Stationary Times Series," *Annals of Mathematical Statistics*, 28(2), 329–348.

POINCARÉ, H. (1908) : *Thermodynamique*. Gauthier-Villars.

POWELL, J. (1983) : "The Asymptotic Normality of Two-stage Least Absolute Deviations Estimators," *Econometrica*, 51(5), 1569–1576.

——— (1984) : "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25(3), 303–325.

——— (1986) : "Censored Regression Quantiles," *Journal of Econometrics*, 32(1), 143–155.

——— (1994) : "Estimation of Semiparametric Models," in Engle and McFadden (1994), chap. 41, pp. 2443–2521.

PRATT, J. W., AND J. D. GIBBONS (1981) : *Concepts of Nonparametric Theory*. Springer-Verlag, New York.

PRESS, W., S. TEUKOLSKY, W. VETTERLING, AND B. FLANNERY (2002) : *Numerical Recipes in Fortran 90*. Press Syndicate of the University of Cambridge, USA.

SAKATA, S. (2001) : "Instrumental Variable Estimation Based on Mean Absolute Deviation," Discussion paper, Michigan - Center for research on economics and social theory, First version 1998, revised 2001.

SCHWEDER, T., AND N. L. HJORT (2002) : "Confidence and Likelihood," *Scandinavian Journal of Statistics*, 29(2), 309–332.

SO, B. (1994) : "A sharp Cramér-Rao Type Lower-Bound for Median-Unbiased Estimators," *Journal of the Korean Statistical Society*, 23(1).

STAIGER, D., AND J. STOCK (1997) : "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65(3), 557–586.

STOCK, J., AND J. WRIGHT (2000) : "GMM with Weak Identification," *Econometrica*, 68(5), 1055–1096.

SUNG, N. K., G. STANGENHAUS, AND H. DAVID (1990) : "A Cramér-Rao Analogue for Median-Unbiased Estimators," *Trabajos de Estadistica*, 5, 83–94.

TIPPETT, L. (1931) : *The Methods of Statistics*. Williams and Norgate, London.

VALÉRY, P., AND J.-M. DUFOUR (2004) : "Finite-Sample and Simulation-Based Tests for Stochastic Volatility Model," Discussion paper, Université de Montréal, CIRANO, CIREQ.

WANG, J., AND E. ZIVOT (1998) : "Inference on Structural Parameters in Instrumental Variables Regression with weak instruments," *Econometrica*, 66(6), 1389–1404.

WEISS, A. (1990) : "Least Absolute Error Estimation in the Presence of Serial Correlation," *Journal of Econometrics*, 44(1), 127–158.

——— (1991) : "Estimating Nonlinear Dynamic Models Using Least Absolute Error Estimation," *Econometric Theory*, 7(1), 46–68.

WHITE, H. (1980) : "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48(4), 817–838.

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

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 Presence of Heteroskedasticity of Unknown Form," *Econometric Theory*, 17, 765–784.

□

Conclusion générale

Nous avons développé dans cette thèse, un système d'inférence exacte pour des modèles semi-paramétriques de régression et des modèles structurels avec instruments. Ce système, qui s'appuie sur des transformations par les signes et la technique des tests de Monte Carlo, donne des résultats valides, quelle que soit la taille de l'échantillon, pour des erreurs hétéroscédastiques de forme très générale. L'inférence reste par ailleurs asymptotiquement valide en présence de dépendance linéaire. La thèse se compose de trois essais.

Dans le premier essai, nous avons étudié des statistiques de signes pivotales, et nous avons construit des tests simultanés pour le vecteur de paramètres d'une régression linéaire sur la médiane. En échantillon fini, le niveau de ces tests égale le niveau nominal si les signes des erreurs satisfont une certaine condition d'invariance (« mediangale »). Les tests restent asymptotiquement valides en présence de processus plus généraux comme, par exemple, des ARMA stationnaires. Nous avons ensuite utilisé les méthodes d'inversion et de projection pour construire des régions de confiance et des tests d'une hypothèse générale possiblement non linéaire. Les études par simulation suggèrent que, dans des échantillons de petite taille, les tests et les régions de confiance que nous proposons sont plus fiables que les méthodes habituelles (moindres carrés, LAD) dès lors que les données sont hétérogènes. Ceci reste vrai même quand ces méthodes habituelles sont corrigées par un « bootstrap ». La procédure proposée s'avère aussi préférable à une version asymptotique lorsqu'elle n'est qu'asymptotiquement justifiée. Prenons l'exemple de données peu nombreuses et linéairement dépendantes. L'approche à distance finie n'est alors qu'asymptotiquement justifiée. Pourtant, elle permet de prendre en compte la distortion due à la petite taille de l'échantillon, ce que ne font pas les approches asymptotiques habituelles. Nous avons présenté deux exemples d'application. Le premier teste la théorie de convergence β entre les niveaux de production des états américains entre 1880 et 1988. Le second teste la présence d'une tendance dans l'indice « Standard and Poor's Composite Price » entre 1928 et 1987, ainsi que pour diverses sous périodes.

Dans le deuxième essai, nous avons developpé plusieurs outils d'inférence à distance finie facilement utilisables dans le système précédent. La fonction p-value (ainsi que ses versions individuelles qui s'obtiennent par projection) résume graphiquement l'inférence disponible sur un paramètre. Elle mesure le degré de confiance que l'on peut avoir dans une valeur donnée du paramètre et permet facilement d'étendre les notions de distribution de confiance à un vecteur de paramètres. Fonction p-value et distribution de confiance s'appuient sur une réinterprétation des distributions fiducielles de Fisher. Elles fournissent, en un sens, un équivalent fréquentiste aux distributions *a posteriori* bayésiennes. Des tests, nous sommes ensuite passés à l'estimation. Nous avons introduit un estimateur de signe grâce au principe de Hodges-Lehmann. Il s'agit de la valeur du paramètre associée à la plus grande p-value ; ou encore la valeur la moins rejetée quand le niveau des tests augmente ; autrement dit, la valeur ayant le plus fort degré de confiance. Cet estimateur ne s'appuie pas sur des considérations asymptotiques contrairement au principe d'analogie. Toutefois, maximiser la p-value équivaut parfois (en probabilité) à une méthode des moments classique dans laquelle les conditions de moments font intervenir les signes. Nous avons étudié les propriétés de ces estimateurs. Ils présentent plusieurs formes d'invariance et sont sans biais pour la médiane lorsque les erreurs sont symétriques. Les conditions de convergence et de normalité asymptotique des estimateurs de signes sont aussi plus faibles que celles requises par l'estimateur LAD. En particulier, la variance des erreurs peut ne pas être finie. D'après nos simulations, les estimateurs de signes ont de bonnes propriétés dans les cas habituels et sont plus fiables que les méthodes de moindres carrés ou que le LAD quand les données sont très hétérogènes. Malgré le fait qu'ils font intervenir des méthodes numériques, nous conseillons de combiner les estimateurs de signe à la méthode d'inférence présentée dans le premier essai lorsque les données sont peu nombreuses ou lorsqu'elles semblent très hétérogènes.

Le troisième essai a porté sur les modèles structurels et les modèles non-linéaires en présence d'instruments. Nous avons développé une procédure d'inférence exacte qui est aussi robuste au degré d'identification du paramètre structurel. Celle-ci s'appuie sur une condition de validité des instruments vis-à-vis des signes de l'erreur structurelle. La distribution des signes est alors pivotale et facilement simulable, et l'on peut construire des tests de Monte Carlo et des régions de confiance par inversion. Notre approche est dans l'esprit d'Anderson et Rubin [Anderson et Rubin (1949)]. Les statistiques IV que nous étudions se déduisent de régressions artificielles des signes des résidus sur des instruments dits « auxiliaires ». Ces instruments n'entrent pas dans la spécification économique du modèle, ils ne servent qu'à calculer la statistique. Les statistiques IV correspondent aussi à des combinaisons de conditions d'orthogonalité qui font intervenir les signes. Nous avons considéré deux types de combinaisons : la forme quadratique habituelle de laquelle découlent des statistiques de type GMM ou Fisher et qui permet d'associer un estimateur, ainsi qu'une approche de type Tippett qui combine les p-values de chaque condition d'orthogonalité testée séparement. Cette dernière approche semble donner de meilleurs résultats que la précédente en cas de faible identification du paramètre. Les tests issus de ces statistiques sont exacts et robustes aux problèmes d'identification. Nous nous sommes aussi demandé quels instruments inclure dans le modèle en cas de suridentification. Nous avons présenté deux concepts d'optimalité des instruments selon les propriétés de puissance des tests qui leur sont associés. Enfin, nous avons présenté des estimateurs. Ceux-ci, comme tout estimateur, ne doivent être utilisés que lorsque le paramètre est identifié. Ils sont convergents et asymptotiquement normaux sous des conditions plus faibles que celles requises dans la théorie des doubles moindres carrés et de l'estimateur « Two-Stage Least Absolute Deviations, TSLAD ». Ces propriétés restent, entre autres, valables si les erreurs présentent des queues de distributions épaisses. Les simulations suggèrent que les tests de signes sont plus performants que les tests usuels (y compris ceux qui sont robustes à la présence d'instruments faibles ou à un manque d'identification) en échantillon fini, quand les données sont hétérogènes, hétéroscédastiques ou lorsque la variable endogène influe sur la distribution de l'erreur structurelle sans en affecter le signe. Enfin, comme exemple d'application, nous sommes revenus sur le problème des rendements de l'éducation de Angrist et Krueger (1991).

Cette thèse a fourni l'occasion d'insister sur l'intérêt des approches à distance finie dans divers modèles économétriques. La transformation par les signes permet d'étendre ces approches à des modèles non ou semi-paramétriques sous des hypothèses distributionnelles très générales.

Plusieurs extensions à ce travail sont envisageables. Les méthodes proposées sont tout d'abord aisémment adaptables à d'autres quantiles que la médiane [voir Koenker et Bassett (1978) qui introduisent les régressions quantile]. La théorie ne changera pas ou très peu. En revanche, le champ d'application pratique en sera fortement étendu. Ensuite, le lien entre p-value et identification pourrait être exploité pour construire des tests de spécification. Enfin, les statistiques de signes étudiées sont souvent des formes quadratiques associées aux tests localement optimaux. Développer d'autres classes de statistiques de signes constitue une extension prometteuse à ce travail, ce que suggère les performances des tests obtenus à partir de la statistique de type Tippett. Il pourrait s'avérer ainsi judicieux de combiner des statistiques de test point-optimal en ayant recours à des inégalités de type Bonferroni ou à des méthodes adaptatives. De telles approches pourraient permettre des gains supplémentaires en puissance selon les cas étudiés.