Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Econometrics 133 (2006) 443–477

JOURNAL OF Econometrics

www.elsevier.com/locate/jeconom

Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics

Jean-Marie Dufour*

Département de sciences économiques, Université de Montréal, C.P. 6128 succursale Centre-ville, Montréal, Québec, Canada H3C 3J7

Available online 30 August 2005

Abstract

The technique of Monte Carlo (MC) tests [Dwass (1957, Annals of Mathematical Statistics 28, 181–187); Barnard (1963, Journal of the Royal Statistical Society, Series B 25, 294)] provides a simple method for building exact tests from statistics whose finite sample distribution is intractable but can be simulated (when no nuisance parameter is involved). We extend this method in two ways: first, by allowing for MC tests based on exchangeable possibly discrete test statistics; second, by generalizing it to statistics whose null distribution involves nuisance parameters [maximized MC (MMC) tests]. Simplified asymptotically justified versions of the MMC method are also proposed: these provide a simple way of improving standard asymptotics and dealing with nonstandard asymptotics.

JEL classification: C12; C15; C2; C52; C22

Keywords: Monte Carlo test; Maximized Monte Carlo test; Finite-sample test; Exact test; Nuisance parameter; Bounds; Bootstrap; Parametric bootstrap; Simulated annealing; Asymptotics; Nonstandard asymptotic distribution

*Tel.: +1 514 343 2400; fax: +1 514 343 5831.
E-mail address: jean.marie.dufour@umontreal.ca.
URL: http://www.fas.umontreal.ca/SCECO/Dufour.

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2005.06.007

1. Introduction

During the last 25 years, the development of faster and cheaper computers has made Monte Carlo techniques more affordable and attractive in statistical analysis. In particular, such techniques may now be used routinely for data analysis. Important developments in this area include the use of bootstrap techniques for improving standard asymptotic approximations (for reviews, see Efron, 1982; Beran and Ducharme, 1991; Efron and Tibshirani, 1993; Hall, 1992; Jeong and Maddala, 1993; Vinod, 1993; Shao and Tu, 1995; Davison and Hinkley, 1997; Chernick, 1999; Horowitz, 1997) and techniques where estimators and forecasts are obtained from criteria evaluated by simulation (see McFadden, 1989; Mariano and Brown, 1993; Hajivassiliou, 1993; Keane, 1993; Gouriéroux and Monfort, 1996; Gallant and Tauchen, 1996).

With respect to tests and confidence sets, these techniques only have asymptotic justifications and do not yield inferences that are provably valid (in the sense of correct levels) in finite samples. Here, it is of interest to note that the use of simulation in the execution of tests was suggested much earlier than recent bootstrap and simulation-based techniques. For example, randomized tests have been proposed long ago as a way of obtaining tests with any given level from statistics with discrete distributions (e.g., sign and rank tests); see Lehmann (1986). A second interesting possibility is the technique of Monte Carlo tests originally suggested by Dwass (1957) for implementing permutation tests and later extended by Barnard (1963), Hope (1968) and Birnbaum (1974). This technique has the great attraction of providing *exact* (randomized) tests based on any statistic whose finite-sample distribution may be intractable but can be simulated. The validity of the tests so obtained does not depend at all on the number of replications made (which can be small). Only the power of the procedure is influenced by the number of replications, but the power gains associated with lengthy simulations are typically rather small. For further discussion of Monte Carlo tests, see Besag and Diggle (1977), Dufour and Kiviet (1996, 1998), Edgington (1980), Edwards (1985), Edwards and Berry (1987), Foutz (1980), Jöckel (1986), Kiviet and Dufour (1997), Marriott (1979) and Ripley (1981).

An important limitation of the technique of Monte Carlo tests is the fact that one needs to have a statistic whose distribution does not depend on nuisance parameters. This obviously limits considerably its applicability. The main objective of this paper is to extend the technique of Monte Carlo tests in order to allow for the presence of nuisance parameters in the null distribution of the test statistic.

In Section 2, we summarize and extend results on Monte Carlo (MC) tests when the null distribution of a test statistic does not involve nuisance parameters. In particular, we put them in a form that will make their extension to cases with nuisance parameters easy and intuitive, and we generalize them by allowing for MC tests based on exchangeable (possibly nonindependent) replications and statistics with discrete distributions. These generalizations allow, in particular, for various nonparametric tests (e.g., permutation tests) as well as test statistics where certain parameters are themselves evaluated by simulation. We deal with possibly discrete (or mixtures of continuous and discrete distributions) by exploiting Hájek's (1969) method of randomized ranks for breaking ties in rank tests, which is both simple to implement and allows one to easily deal with exchangeable [as opposed to independent and identically distributed (i.i.d.)] simulations. On the problem of discrete distributions, it is also of interest that the method proposed by Jöckel (1986) was derived under the assumption of i.i.d. MC replications.

In Section 3, we study how the power of Monte Carlo tests is related to the number of replications used and the sensitivity of the conclusions to the randomized nature of the procedure. In particular, given the observed (randomized) p-value of the Monte Carlo test, we see that the probability of an eventual reversal of the conclusion of the procedure (rejection or acceptance at a given level, e.g. 5%) can easily be computed.

In Section 4, we present the extension to statistics whose null distribution depends on nuisance parameters. This procedure is based on considering a simulated p-value function which depends on nuisance parameters (under the null hypothesis). We show that maximizing the latter with respect to the nuisance parameters yields a test with provably exact level, irrespective of the sample size and the number replications used. For this reason, we call the latter maximized Monte Carlo (MMC) tests. As one would expect for a statistic whose distribution depends on unknown nuisance parameters, the probability of type I error for a MMC test can be lower (but not higher) than the level of the test, so the procedure can be conservative. We also discuss how this maximization can be achieved in practice, e.g. through *simulated annealing* techniques.

In the two next sections, we discuss simplified (asymptotically justified) approximate versions of the proposed procedures, which involve the use of consistent set or point estimates of model parameters. In Section 5, we suggest a method [the consistent set estimate MMC method (CSEMMC)] which is applicable when a consistent set estimator of the nuisance parameters [e.g., a random subset of the parameter space whose probability of covering the nuisance parameters converges to one as the sample size goes to infinity] is available. The approach proposed involves maximizing the simulated p-value function over the consistent set estimate, as opposed to the full nuisance parameter space. This procedure may thus be computationally much less costly. Using a consistent set estimator (or confidence set), as opposed to a point estimate, to deal with nuisance parameters is especially useful because it allows one to obtain asymptotically valid tests even when the test statistic does not converge in distribution or when the asymptotic distribution depends on nuisance parameters possibly in a discontinuous way. Consequently, there is no need to study the asymptotic distribution of the test statistic considered or even to establish its existence.[1] This consistent set estimator MMC method

[1]A case where the distribution of a test statistic does not converge in distribution is the one where the associated sequence of distribution functions has several accumulation points, allowing different subsequences to have different limiting distributions.

(CSEMMC) may be viewed as an asymptotic Monte Carlo extension of finite-sample two-stage procedures proposed in Dufour (1990), Dufour and Kiviet (1996, 1998), Campbell and Dufour (1997), and Dufour et al. (1998). These features may be contrasted with those of bootstrap methods which can fail to provide asymptotically valid tests when the test statistic simulated has an asymptotic distribution involving nuisance parameters, especially if the asymptotic distribution has discontinuities with respect to the nuisance parameters (see Athreya, 1987; Basawa et al., 1991; Sriram, 1994; Andrews, 2000; Benkwitz et al., 2000; Inoue and Kilian, 2002, 2003).

In Section 6, we consider the simplest form of a Monte Carlo test with nuisance parameters, i.e. the one where the consistent set estimate has been replaced by a consistent point estimate. In other words, the distribution of the test statistic is simulated after replacing the nuisance parameters by a consistent point estimate. Such a procedure can be interpreted as a parametric bootstrap test based on the percentile method (see Efron and Tibshirani, 1993, Chapter 16; Hall, 1992). The term "parametric" may however be misleading here, because such MC tests can be applied as well to nonparametric (distribution-free) test statistics. We give general conditions under which a Monte Carlo test obtained after replacing an unknown nuisance parameter yield an asymptotically valid test in cases where the limit distribution of the test statistic involves nuisance parameters. Following the general spirit of Monte Carlo testing and in contrast with typical bootstrap arguments, the proofs take the number of Monte Carlo simulations as fixed (possibly very small, such as 19 to obtain a test with level 0.05). As in standard bootstrap arguments, the conditions considered involve a smooth (continuous) dependence of the asymptotic distribution upon the nuisance parameters. It is, however, important to note that these conditions are more restrictive and more difficult to check than those under which CSEMMC procedures would be applicable. We conclude in Section 7.

2. Monte Carlo tests without nuisance parameters

Let us consider a family of probability spaces $\{(\mathscr{X}, \mathscr{A}_{\mathscr{X}}, \mathsf{P}_\theta) : \theta \in \Omega\}$, where $\mathscr{X}$ is a sample space, $\mathscr{A}_{\mathscr{X}}$ a σ-algebra of subsets of $\mathscr{X}$, and Ω a parameter space (possibly infinite dimensional). Let also $S \equiv S(\omega)$, $\omega \in \mathscr{X}$, be a real-valued $\mathscr{A}_{\mathscr{X}}$ -measurable function whose distribution is determined by P_{θ_0}—i.e., θ_0 is the "true" parameter vector. We wish to test the hypothesis

$$H_0 : \theta_0 \in \Omega_0, \tag{2.1}$$

where Ω_0 is a nonempty subset of Ω, using a critical region of the form $\{S \geqslant c\}$. Although, in general, the distribution of S under H_0 depends on the unknown value of θ_0, we shall assume in this section that this distribution does not depend on (unknown) nuisance parameters, so that we can write

$$\mathsf{P}_\theta[S \leqslant x] = F(x) \quad \text{for all } \theta \in \Omega_0, \tag{2.2}$$

where $F(x)$ is the unique distribution that S can have under H_0. In view of this assumption, we shall—until further notice—compute probabilities under the (unique) $\mathsf{P} \equiv \mathsf{P}_{\theta_0}$ when $\theta_0 \in \Omega_0$. The constant c is chosen so that

$$\mathsf{P}[S \geqslant c] = 1 - F(c) + \mathsf{P}[S = c] \leqslant \alpha, \tag{2.3}$$

where α is the desired level of the test $(0 < \alpha < 1)$. Note that the critical region $S \geqslant c$ can also be put in two useful alternative forms, which are equivalent to $S \geqslant c$ with probability one (i.e., they can differ from the critical region $S \geqslant c$ only on a set of zero probability):

$$G(S) \leqslant G(c), \tag{2.4}$$

$$S \geqslant F^{-1}[(F(c) - \mathsf{P}[S = c])^+] = F^{-1}[(1 - G(c))^+], \tag{2.5}$$

where

$$G(x) = \mathsf{P}[S \geqslant x] = 1 - F(x) + \mathsf{P}[S = x] \tag{2.6}$$

is the "tail-area" or "p-value" function associated with F, and F^{-1} is the quantile function of F, with the conventions

$$F^{-1}(q^+) = \lim_{\varepsilon \downarrow 0} F^{-1}(q + \varepsilon) = \inf\{F^{-1}(q_0) : q_0 > q\}, \quad 0 \leqslant q \leqslant 1,$$

$F^{-1}(1^+) = \infty$ and $F^{-1}(0^+) = F^{-1}(0)$. For any probability distribution function $F(x)$, the quantile function $F^{-1}(q)$ is defined as follows:

$$\begin{aligned} F^{-1}(q) &= \inf\{x : F(x) \geqslant q\} \quad \text{if } 0 < q < 1, \\ &= \inf\{x : F(x) > 0\} \quad \text{if } q = 0, \\ &= \sup\{x : F(x) < 1\} \quad \text{if } q = 1; \end{aligned} \tag{2.7}$$

see Reiss (1989, p. 13). In general, $F^{-1}(q)$ takes its values in the extended real numbers $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ and, for coherence, we set $F(-\infty) = 0$ and $F(\infty) = 1$. Using (2.7), it is easy to see that

$$F^{-1}[(F(c) - \mathsf{P}[S = c])^+] = c,$$

when: $0 < F(c) < 1$ and either $\mathsf{P}(S = c) > 0$ or $F(x)$ is continuous and strictly monotonic in an open neighborhood containing c. However, formulation (2.5) remains valid in all cases.

2.1. Monte Carlo tests based on statistics with continuous distributions

Consider now a situation where the distribution of S under H_0 may not be easy to compute analytically but can be simulated. Let $S_1, \ldots, S_N$ be a sample of identically distributed real random variables with the same distribution as S. Typically, it is assumed that $S_1, \ldots, S_N$ are also independent. However, we will observe that the exchangeability of $S_1, \ldots, S_N$ is sufficient for most of the results presented below.[2]

[2]The elements of a random vector $(S_1, S_2, \ldots, S_N)'$ are exchangeable (or P-exchangeable) if $(S_{r_1}, S_{r_2}, \ldots, S_{r_N})' \sim (S_1, S_2, \ldots, S_N)'$ for any permutation $(r_1, r_2, \ldots, r_N)$ of the integers $(1, 2, ..., N)$ under the relevant probability measure P.

This can accommodate a wide array of situations, where the simulated statistics are not independent because they involve common (conditioning) variables, such as: statistics obtained by permuting randomly a given set of observations (permutation tests), tests which are simulated conditionally on a common set of initial values (e.g., in time series models), common regressors or a common subsample [see the Anderson–Rubin-type split-sample test described in Dufour and Jasiak (2001)], tests that depend on a common independent simulation [e.g., tests based on a criterion evaluated by a preliminary simulation, such as the simulated method of moments or indirect inference; see Gouriéroux and Monfort (1996)], etc.

The technique of MC tests provides a simple method allowing one to replace the theoretical distribution $F(x)$ by its sample analogue based on $S_1, \ldots, S_N$:

$$\hat{F}_N(x) \equiv \hat{F}_N[x; S(N)] = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(S_i \leqslant x), \tag{2.8}$$

where $S(N) = (S_1, \ldots, S_N)'$ and $\mathbf{1}(C)$ is the indicator function associated with the condition C:

$$\begin{aligned} \mathbf{1}(C) &= 1 \quad \text{if condition } C \text{ holds,} \\ &= 0 \quad \text{otherwise.} \end{aligned} \tag{2.9}$$

In the latter notation, C may also be replaced by an event $A \in \mathcal{A}_{\mathcal{X}}$, in which case $\mathbf{1}(A) \equiv \mathbf{1}(\omega \in A)$ where ω is the element drawn from the sample space.

We also consider the corresponding sample *tail area* (or *survival*) function:

$$\hat{G}_N[x; S(N)] = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(S_i \geqslant x). \tag{2.10}$$

The sample distribution function is related to the ranks $R_1, \ldots, R_N$ of the variables $S_1, \ldots, S_N$ (when put in ascending order) by the expression:

$$R_j = N\hat{F}_N[S_j; S(N)] = \sum_{i=1}^{N} \mathbf{1}(S_i \leqslant S_j), \quad j = 1, \ldots, N. \tag{2.11}$$

The central property we shall exploit here is the following: to obtain critical values or compute p-values, the "theoretical" null distribution $F(x)$ can be replaced by its simulation-based "estimate" $\hat{F}_N(x)$ in a way that will preserve the level of the test in *finite samples*, *irrespective of the number N of replications used*. For continuous distributions, this property is expressed by Proposition 2.2 below, which is easily proved by using the following simple lemma.

Lemma 2.1 (*Distribution of ranks when ties have zero probability*). *Let* $(y_1, \ldots, y_N)'$ *be a* $N \times 1$ *vector of* P*-exchangeable real random variables such that*

$$\mathsf{P}(y_i = y_j) = 0 \quad \textit{for } i \neq j, \;\; i,j = 1, \ldots, N, \tag{2.12}$$

and let $R_j = \sum_{i=1}^{N} \mathbf{1}(y_i \leqslant y_j)$ *be the rank of* y_j *when* $y_1, \ldots, y_N$ *are ranked in nondecreasing order* ($j = 1, \ldots, N$). *Then, for* $j = 1, \ldots, N$,

$$\mathsf{P}(R_j/N \leqslant x) = I[xN]/N \quad \text{for } 0 \leqslant x \leqslant 1, \tag{2.13}$$

$$\begin{aligned} \mathsf{P}(R_j/N \geqslant x) &= 1 \quad \text{if } x \leqslant 0, \\ &= (I[(1-x)N] + 1)/N \quad \text{if } 0 < x \leqslant 1, \\ &= 0 \quad \text{if } x > 1, \end{aligned} \tag{2.14}$$

where $I[x]$ *is the largest integer less than or equal to* x.

Note that we use the symbol $I[x]$ rather than the common notation $[x]$ to represent the integer part of a number x, because we heavily use brackets elsewhere in the paper, so that the notation $[\cdot]$ could lead to confusions. It is clear that condition (2.12) is satisfied whenever the variables $y_1, \ldots, y_N$ are independent with continuous distribution functions (see Hájek, 1969, pp. 20–21), or when the vector $(y_1, \ldots, y_N)'$ has an absolutely continuous distribution (with respect to the Lebesgue measure on $\mathbb{R}^N$).

Proposition 2.2 (*Validity of Monte Carlo tests when ties have zero probability*). *Let* $(S_0, S_1, \ldots, S_N)'$ *be a* $(N+1) \times 1$ *vector of exchangeable real random variables such that*

$$\mathsf{P}(S_i = S_j) = 0 \quad \text{for } i \neq j,\ i, j = 0, 1, \ldots, N, \tag{2.15}$$

let $\hat{F}_N(x) \equiv \hat{F}_N[x; S(N)]$, $\hat{G}_N(x) = \hat{G}_N[x; S(N)]$ *and* $\hat{F}_N^{-1}(x)$ *be defined as in* (2.7)–(2.10), *and set*

$$\hat{p}_N(x) = \frac{N\hat{G}_N(x) + 1}{N+1}. \tag{2.16}$$

Then,

$$\mathsf{P}[\hat{G}_N(S_0) \leqslant \alpha_1] = \mathsf{P}[\hat{F}_N(S_0) \geqslant 1 - \alpha_1] = \frac{I[\alpha_1 N] + 1}{N+1} \quad \text{for } 0 \leqslant \alpha_1 \leqslant 1, \tag{2.17}$$

$$\mathsf{P}[S_0 \geqslant \hat{F}_N^{-1}(1 - \alpha_1)] = \frac{I[\alpha_1 N] + 1}{N+1} \quad \text{for } 0 < \alpha_1 < 1, \tag{2.18}$$

and

$$\mathsf{P}[\hat{p}_N(S_0) \leqslant \alpha] = \frac{I[\alpha(N+1)]}{N+1} \quad \text{for } 0 \leqslant \alpha \leqslant 1. \tag{2.19}$$

The latter proposition can be used as follows: choose α_1 and N so that

$$\alpha = \frac{I[\alpha_1 N] + 1}{N+1} \tag{2.20}$$

is the desired significance level. Provided N is reasonably large, α_1 will be very close to α; in particular, if $\alpha(N+1)$ is an integer, we can take $\alpha_1 = \alpha - ((1-\alpha)/N)$, in which case we see easily that the critical region $\hat{G}_N(S_0) \leqslant \alpha_1$ is equivalent to $\hat{G}_N(S_0) < \alpha$. Further, for $0 < \alpha < 1$, the randomized critical region $S_0 \geqslant \hat{F}_N^{-1}(1 - \alpha_1)$ has the same level (α) as the nonrandomized critical region $S_0 \geqslant F^{-1}(1-\alpha)$, or equivalently the critical regions $\hat{p}_N(S_0) \leqslant \alpha$ and $\hat{G}_N(S_0) \leqslant \alpha_1$ have the same level as the critical region $G(S_0) \equiv 1 - F(S_0) \leqslant \alpha$.

2.2. Monte Carlo tests based on general statistics

Assumption (2.15), which states that ties have zero probability, plays an important role in proving Proposition 2.2. However, it is possible to prove analogous results for general sequences of exchangeable random variables (which may exhibit ties with positive probability), provided we consider a properly randomized empirical distribution function. For this purpose, we introduce randomized ranks which are obtained like ordinary ranks except that ties are "broken" according to a uniform distribution. More precisely, let us associate with each variable S_j, $j = 1, \ldots, N$, a random variable U_j, $j = 1, \ldots, N$ such that

$$U_1, \ldots, U_N \overset{\text{i.i.d.}}{\sim} U(0,1), \tag{2.21}$$

$U(N) = (U_1, \ldots, U_N)'$ is independent of $S(N) = (S_1, \ldots, S_N)'$ where $U(0,1)$ is the uniform distribution on the interval $(0, 1)$. Then, we consider the pairs

$$Z_j = (S_j, U_j), \quad j = 1, \ldots, N, \tag{2.22}$$

which are ordered according to the lexicographic order:

$$(S_i, U_i) \leqslant (S_j, U_j) \Longleftrightarrow \{S_i < S_j \text{ or } (S_i = S_j \text{ and } U_i \leqslant U_j)\}. \tag{2.23}$$

Using the indicator

$$\mathbf{1}[(x_1, u_1) \leqslant (x_2, u_2)] = \mathbf{1}(x_1 < x_2) + \delta(x_1 - x_2)\mathbf{1}(u_1 \leqslant u_2), \tag{2.24}$$

$S_1, \ldots, S_N$ are then ordered like the pairs $Z_1, \ldots, Z_N$ according to (2.23), which yield "randomized ranks":

$$\tilde{R}_j[S(N), U(N)] = \sum_{i=1}^{N} \mathbf{1}[(S_i, U_i) \leqslant (S_j, U_j)], \tag{2.25}$$

$j = 1, \ldots, N$. By the continuity of the uniform distribution, the ranks $\tilde{R}_j = \tilde{R}_j[S(N), U(N)]$, $j = 1, \ldots, N$, are all distinct with probability 1, so that the randomized rank vector $(\tilde{R}_1, \tilde{R}_2, \ldots, \tilde{R}_N)'$ is a permutation of $(1, 2, \ldots, N)'$ with probability 1. Furthermore when $S_j \neq S_i$ for all $j \neq i$, we have $\tilde{R}_j = R_j$: if (2.15) holds, then $\tilde{R}_j = R_j$, $j = 1, \ldots, N$, with probability 1 [where R_j is defined in (2.11)]. We can now state the following extension of Lemma 2.1.

Lemma 2.3 (*Distribution of randomized ranks*). *Let $y(N) = (y_1, \ldots, y_N)'$ be a $N \times 1$ vector of exchangeable real random variables and let $\tilde{R}_j = \tilde{R}_j[y(N), U(N)]$be defined as in* (2.25) *where $U(N) = (U_1, \ldots, U_N)'$ is a vector of i.i.d. $U(0,1)$ variables independent of $y(N)$. Then, for $j = 1, \ldots, N$,*

$$\mathsf{P}(\tilde{R}_j/N \leqslant x) = I[xN]/N \quad \text{for } 0 \leqslant x \leqslant 1, \tag{2.26}$$

$$\begin{aligned}\mathsf{P}(\tilde{R}_j/N \geqslant x) &= 1 \quad \textit{if } x \leqslant 0, \\ &= (I[(1-x)N]+1)/N \quad \textit{if } 0 < x \leqslant 1, \\ &= 0 \quad \textit{if } x > 1.\end{aligned} \tag{2.27}$$

To the above randomized rankings, it is natural to associate the following randomized empirical (pseudo-)distribution function:

$$\begin{aligned}\tilde{F}_N[x; U_0, S(N), U(N)] &= \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[(S_i, U_i) \leqslant (x, U_0)] \\ &= 1 - \hat{G}_N[x; S(N)] + T_N[x; U_0, S(N), U(N)],\end{aligned} \tag{2.28}$$

where U_0 is a $U(0,1)$ random variable independent of $S(N)$ and $U(N)$,

$$T_N[x; U_0, S(N), U(N)] = \frac{1}{N}\sum_{i=1}^{N} \delta(S_i - x)\mathbf{1}(U_i \leqslant U_0) = \frac{1}{N}\sum_{i \in E_N(x)} \mathbf{1}(U_i \leqslant U_0) \tag{2.29}$$

and $E_N(x) = \{i : S_i = x,\ 1 \leqslant i \leqslant N\}$. The function $\tilde{F}_N[x; \cdot]$ retains all the properties of a probability distribution function, except for the fact that it may not be right continuous at some of its jump points (where it may take values between its right and left limits). We can also define the corresponding tail-area function:

$$\begin{aligned}\tilde{G}_N[x; U_0, S(N), U(N)] &= \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[(S_i, U_i) \geqslant (x, U_0)] \\ &= 1 - \hat{F}_N[x; S(N)] + \bar{T}_N[x; U_0, S(N), U(N)],\end{aligned} \tag{2.30}$$

$$\bar{T}_N[x; U_0, S(N), U(N)] = \frac{1}{N}\sum_{i=1}^{N} \delta(S_i - x)\mathbf{1}(U_i \geqslant U_0) = \frac{1}{N}\sum_{i \in E_N(x)} \mathbf{1}(U_i \geqslant U_0). \tag{2.31}$$

From (2.28)–(2.31), we see that the following inequalities must hold:

$$1 - \hat{G}_N[x; S(N)] \leqslant \tilde{F}_N[x; U_0, S(N), U(N)] \leqslant \hat{F}_N[x; S(N)], \tag{2.32}$$

$$1 - \hat{F}_N[x; S(N)] \leqslant \tilde{G}_N[x; U_0, S(N), U(N)] \leqslant \hat{G}_N[x; S(N)]. \tag{2.33}$$

When no element of $S(N)$ is equal to x [i.e., when $E_N(x)$ is empty], we have:

$$\begin{aligned}\tilde{G}_N[x; U_0, S(N), U(N)] &= \hat{G}_N[x; S(N)] = 1 - \hat{F}_N[x; S(N)] \\ &= 1 - \tilde{F}_N[x; U_0, S(N), U(N)].\end{aligned} \tag{2.34}$$

Using the above observations, it is then easy to establish the following proposition.

Proposition 2.4 (*Validity of Monte Carlo tests for general statistics*). *Let* $(S_0, S_1, \ldots, S_N)'$ *be a* $(N+1) \times 1$ *vector of exchangeable real random variables, let* $(U_0, U_1, \ldots, U_N)'$ *be a* $(N+1) \times 1$ *vector of i.i.d.* $U(0,1)$ *random variables independent of* $(S_0, S_1, \ldots, S_N)'$, *let* $\hat{F}_N(x) \equiv \hat{F}_N[x; S(N)]$, $\hat{G}_N(x) \equiv \hat{G}_N[x; S(N)]$, $\tilde{F}_N(x) \equiv \tilde{F}_N[x; U_0, S(N), U(N)]$ *and* $\tilde{G}_N(x) \equiv \tilde{G}_N[x; U_0, S(N), U(N)]$ *be defined as in* (2.8)–(2.10) *and* (2.28)–(2.30), *with* $S(N) = (S_1, \ldots, S_N)$ *and* $U(N) = (U_1, \ldots, U_N)$, *and let*

$$\tilde{p}_N(x) = \frac{N\tilde{G}_N(x) + 1}{N+1}. \tag{2.35}$$

Then for $0 \leqslant \alpha_1 \leqslant 1$,

$$\begin{aligned} \mathsf{P}[\hat{G}_N(S_0) \leqslant \alpha_1] \leqslant \mathsf{P}[\tilde{G}_N(S_0) \leqslant \alpha_1] &= \mathsf{P}[\tilde{F}_N(S_0) \geqslant 1 - \alpha_1] \\ &= \frac{I[\alpha_1 N] + 1}{N+1} \leqslant \mathsf{P}[\hat{F}_N(S_0) \geqslant 1 - \alpha_1] \end{aligned} \tag{2.36}$$

with $\mathsf{P}[\hat{F}_N(S_0) \geqslant 1 - \alpha_1] = \mathsf{P}[S_0 \geqslant \hat{F}_N^{-1}(1 - \alpha_1)]$ *for* $0 < \alpha_1 < 1$, *and defining* $\hat{p}_N(x)$ *as in* (2.16),

$$\mathsf{P}[\hat{p}_N(S_0) \leqslant \alpha] \leqslant \mathsf{P}[\tilde{p}_N(S_0) \leqslant \alpha] = \frac{I[\alpha(N+1)]}{N+1} \quad \textit{for } 0 \leqslant \alpha \leqslant 1. \tag{2.37}$$

In view of the fact that $\tilde{G}_N(S_0) = \hat{G}_N(S_0)$ with probability one when the zero probability tie condition [i.e. (2.15)] holds, it is straightforward to see that Proposition 2.2 is entailed by Proposition 2.4.

3. Power functions and concordance probabilities

The procedures described above are *randomized* in the sense that the result of the tests depend on auxiliary simulations. This raises the issue of the sensitivity of the results to these simulations. To study this more closely, let us suppose that

$$\begin{aligned} &S_0, S_1, \ldots, S_N \text{ are independent with} \\ &\mathsf{P}(S_i \leqslant x) = F(x), \quad \mathsf{P}(S_i \geqslant x) = G(x), \quad \mathsf{P}(S_i = x) = g(x), \;\; i = 1, \ldots, N, \\ &\mathsf{P}(S_0 \leqslant x) = H(x), \quad \mathsf{P}(S_0 \geqslant x) = K(x). \end{aligned} \tag{3.1}$$

Then, it is easy to see that $N\tilde{G}_N(x)$ follows a binomial distribution $\mathrm{Bi}(N, p)$ with number of trials N and probability of "success" $p = \bar{G}(x, u)$, where

$$\begin{aligned} \bar{G}(x, u) &= \mathsf{P}(\mathbf{1}[(S_i, U_i) \geqslant (x, u)] = 1) = \mathsf{P}(S_i > x) + \mathsf{P}(S_i = x)\mathsf{P}(U_i \geqslant u) \\ &= 1 - F(x) + g(x)(1 - u), \end{aligned} \tag{3.2}$$

and we can compute the conditional probability given (S_0, U_0) of the critical region $\tilde{G}_N(S_0) \leqslant \alpha_1$:

$$\mathsf{P}[\tilde{G}_N(S_0) \leqslant \alpha_1 \mid (S_0, U_0)] = \mathsf{P}\left[\sum_{i=1}^{N} \mathbf{1}[(S_i, U_i) \geqslant (S_0, U_0)] \leqslant I[\alpha_1 N] \mid (S_0, U_0)\right]$$
$$= \sum_{k=0}^{I[\alpha_1 N]} \binom{N}{k} \bar{G}(S_0, U_0)^k [1 - \bar{G}(S_0, U_0)]^{N-k}, \tag{3.3}$$

where $\binom{N}{k} = N!/[k!(N-k)!]$. Similarly, we can also write

$$\mathsf{P}[\hat{G}_N(S_0) \leqslant \alpha_1 \mid S_0] = \sum_{k=0}^{I[\alpha_1 N]} \binom{N}{k} G(S_0)^k (1 - G(S_0))^{N-k}. \tag{3.4}$$

When $F(x)$ is continuous, so that $g(x) = 0$, we have

$$\mathsf{P}[\tilde{G}_N(S_0) \leqslant \alpha_1 \mid (S_0, U_0)] = \mathsf{P}[\hat{G}_N(S_0) \leqslant \alpha_1 \mid S_0]$$
$$= \sum_{k=0}^{I[\alpha_1 N]} \binom{N}{k} [1 - F(S_0)]^k F(S_0)^{N-k}. \tag{3.5}$$

Using (3.3), we can find a closed-form expression for the power of the randomized test $\tilde{G}(S_0) \leqslant \alpha_1$ for any null hypothesis which entails that S_0 has the distribution $F(\cdot)$ against an alternative under which its distribution is $H(\cdot)$:

$$\mathsf{P}[\tilde{G}_N(S_0) \leqslant \alpha_1] = \underset{(S_0, U_0)}{\mathsf{E}} \{\mathsf{P}[\tilde{G}_N(S_0) \leqslant \alpha_1 \mid (S_0, U_0)]\}$$
$$= \sum_{k=0}^{I[\alpha_1 N]} \binom{N}{k} \int \int_0^1 \bar{G}(x, u)^k [1 - \bar{G}(x, u)]^{N-k} \, \mathrm{d}u \, \mathrm{d}H(x). \tag{3.6}$$

Furthermore, when $F(x)$ is continuous everywhere, the latter expression simplifies and we can write:

$$\mathsf{P}[\tilde{G}_N(S_0) \leqslant \alpha_1] = \mathsf{P}[\hat{G}_N(S_0) \leqslant \alpha_1]$$
$$= \sum_{k=0}^{I[\alpha_1 N]} \binom{N}{k} \int [1 - F(x)]^k F(x)^{N-k} \, \mathrm{d}H(x). \tag{3.7}$$

The above formulae will be useful in establishing the validity of simplified asymptotic Monte Carlo tests in the presence of nuisance parameters. They also allow one to compute the probability that the result of the randomized test $\tilde{G}_N(S_0) \leqslant \alpha_1$ be different of the corresponding nonrandomized test $G(S_0) \leqslant \alpha$, where $\alpha \equiv ([N\alpha_1] + 1)/(N + 1)$. For example, let $\hat{\alpha}_0 = G(S_0)$ the "p-value" one would obtain if the function $G(x)$ were easy to compute (the p-value of the "fundamental test"). The latter is generally different from the p-value $\tilde{p}_N(S_0)$ or $\hat{p}_N(S_0)$ obtained from a Monte Carlo test based on $S_1, \ldots, S_N$. An interesting question here is the probability that the Monte Carlo test yields a conclusion different from the one based on $\hat{\alpha}_0$. To study this, we shall consider the test which rejects the null hypothesis H_0 when $\hat{p}_N(S_0) \leqslant \alpha$ under assumptions (3.1).

If $\hat{\alpha}_0 > \alpha$ (in which case H_0 is not rejected at level α by the fundamental test), the probability that H_0 be rejected at level α_0 is

$$\begin{aligned}\mathsf{P}[\hat{p}_N(S_0) \leqslant \alpha_0 \mid S_0] &= \mathsf{P}[N\hat{G}_N(S_0) \leqslant (N+1)\alpha_0 - 1 \mid S_0] \\ &= \mathsf{P}[\mathrm{Bi}(N, \hat{\alpha}_0) \leqslant (N+1)\alpha_0 - 1 \mid S_0] \\ &\leqslant \mathsf{P}[\mathrm{Bi}(N, \alpha) \leqslant (N+1)\alpha_0 - 1] \\ &= \mathsf{P}\left[\frac{\mathrm{Bi}(N,\alpha) - N\alpha}{(N\alpha(1-\alpha))^{1/2}} \leqslant \frac{N(\alpha_0 - \alpha) - (1-\alpha_0)}{(N\alpha(1-\alpha))^{1/2}}\right], \end{aligned} \tag{3.8}$$

where the inequality follows on observing that $\hat{\alpha}_0 > \alpha$ and $\mathrm{Bi}(N, p)$ denotes a binomial random variable with number of trials N and probability of success p. From (3.8), we can bound the probability that a Monte Carlo p-value as low as α_0 be obtained when the fundamental test is not significant at level α. In particular, for $\alpha_0 < \alpha$, this probability decreases as the difference $|\alpha_0 - \alpha|$ and N get larger. It is also interesting to observe that

$$\lim_{N\to\infty} \mathsf{P}[\hat{p}_N(S_0) \leqslant \alpha_0 \mid S_0] = \lim_{N\to\infty} \mathsf{P}\left[\frac{\mathrm{Bi}(N,\alpha) - N\alpha}{(N\alpha(1-\alpha))^{1/2}} \leqslant \frac{N(\alpha_0 - \alpha) - (1-\alpha_0)}{(N\alpha(1-\alpha))^{1/2}}\right] = 0 \tag{3.9}$$

for $\alpha_0 < \alpha$, so that the probability of a discrepancy between the fundamental test and the Monte Carlo test goes to zero as N increases.

Similarly, for $\hat{\alpha}_0 < \alpha$ (in which case H_0 is rejected at level α by the fundamental test), the probability that H_0 not be rejected at level α_0 is

$$\begin{aligned}\mathsf{P}[\hat{p}_N(S_0) > \alpha_0 \mid S_0] &= \mathsf{P}[\mathrm{Bi}(N, \hat{\alpha}_0) > (N+1)\alpha_0 - 1 \mid S_0] \\ &\leqslant \mathsf{P}\left[\frac{\mathrm{Bi}(N,\alpha) - N\alpha}{(N\alpha(1-\alpha))^{1/2}} \leqslant \frac{N(\alpha_0 - \alpha) - (1-\alpha_0)}{(N\alpha(1-\alpha))^{1/2}}\right]\end{aligned} \tag{3.10}$$

hence

$$\lim_{N\to\infty} \mathsf{P}[\hat{p}_N(S_0) \geqslant \alpha_0 \mid S_0] = 0 \quad \text{for } \alpha_0 > \alpha. \tag{3.11}$$

Eq. (3.10) gives an upper bound on the probability of observing a p-value as high as α_0 when the fundamental test is significant at a level lower than α. Again, the probability of a discrepancy between the fundamental test and the Monte Carlo test goes to zero as N increases. The only case where the probability of a discrepancy between the two tests does not go to zero as $N \to \infty$ is when $\hat{\alpha}_0 = \alpha$ (an event with probability zero for statistics with continuous distributions).

The probabilities (3.8) and (3.10) may be computed a posteriori to assess the probability of obtaining p-values as low (or as high) as $\hat{p}_N(S_0)$ when the result of the corresponding fundamental test is actually not significant (or significant) at level α. Note also that similar (although somewhat different) calculations may be used to determine the number N of simulations that will ensure a given probability of concordance between the fundamental and the Monte Carlo test (see Marriott, 1979).

4. Monte Carlo tests with nuisance parameters

We will now study the case where the distribution of the test statistic S depends on nuisance parameters. We consider a family of probability spaces $\{(\mathscr{Z}, \mathscr{A}_{\mathscr{Z}}, \mathsf{P}_\theta) : \theta \in \Omega\}$ and suppose that S is a real-valued $\mathscr{A}_{\mathscr{Z}}$-measurable function whose distribution is determined by P_{θ_0} (i.e., θ_0 is the "true" parameter vector). We wish to test the hypothesis

$$H_0 : \theta_0 \in \Omega_0, \tag{4.1}$$

where Ω_0 is a nonempty subset of Ω. Again we take a critical region of the form $S \geqslant c$, where c is a constant which does not depend on θ. The critical region $S \geqslant c$ has *level* α if and only if

$$\mathsf{P}_\theta[S \geqslant c] \leqslant \alpha, \quad \forall \theta \in \Omega_0, \tag{4.2}$$

or equivalently,

$$\sup_{\theta \in \Omega_0} \mathsf{P}_\theta[S \geqslant c] \leqslant \alpha. \tag{4.3}$$

Furthermore, $S \geqslant c$ has *size* α when

$$\sup_{\theta \in \Omega_0} \mathsf{P}_\theta[S \geqslant c] = \alpha. \tag{4.4}$$

If we define the distribution and p-value functions,

$$F[x \mid \theta] = \mathsf{P}_\theta[S \leqslant x], \quad x \in \bar{\mathbb{R}}, \tag{4.5}$$

$$G[x \mid \theta] = \mathsf{P}_\theta[S \geqslant x], \quad x \in \bar{\mathbb{R}}, \tag{4.6}$$

where $\theta \in \Omega$, it is again easy to see that the critical regions

$$\sup_{\theta \in \Omega_0} G[S \mid \theta] \leqslant \alpha, \tag{4.7}$$

where $\alpha \equiv \sup_{\theta \in \Omega_0} G[c \mid \theta]$, and

$$S \geqslant \sup_{\theta \in \Omega_0} F^{-1}[(1 - G[c \mid \theta])^+ \mid \theta] \equiv \bar{c} \tag{4.8}$$

are equivalent to $S \geqslant c$ in the sense that $c \leqslant \bar{c}$, with equality holding when $F[x \mid \theta]$ is discontinuous at $x = c$ for all $\theta \in \Omega_0$ or both $F[x \mid \theta]$ and $F^{-1}[q \mid \theta]$ are continuous at $x = c$ and $q = F(c)$ respectively for all $\theta \in \Omega_0$, and

$$\sup_{\theta \in \Omega_0} \mathsf{P}_\theta[S \geqslant \bar{c}] \leqslant \sup_{\theta \in \Omega_0} \mathsf{P}_\theta[S \geqslant c] = \sup_{\theta \in \Omega_0} \mathsf{P}_\theta[\sup\{G[S \mid \theta_0] : \theta_0 \in \Omega_0\} \leqslant \alpha]. \tag{4.9}$$

We shall now extend Proposition 2.2 in order to allow for the presence of nuisance parameters. For that purpose, we consider a real random variable S_0 and random vectors of the form

$$S(N, \theta) = (S_1(\theta), \ldots, S_N(\theta))', \quad \theta \in \Omega, \tag{4.10}$$

all defined on a common probability space $(\mathscr{X}, \mathscr{A}_{\mathscr{X}}, \mathsf{P})$, such that

$$\text{the variables } S_0, S_1(\theta_0), \ldots, S_N(\theta_0) \text{ are exchangeable for some } \theta_0 \in \Omega, \\ \text{each one with distribution function } F[x \mid \theta_0]. \tag{4.11}$$

Typically, S_0 will refer to a test statistic computed from observed data when the true parameter vector is θ_0 (i.e., $\theta = \theta_0$), while $S_1(\theta), \ldots, S_N(\theta)$ will refer to independent and identically distributed (*i.i.d.*) replications of the test statistic obtained independently (e.g., by simulation) under the assumption that the parameter vector is θ (i.e., $\mathsf{P}[S_i(\theta) \leqslant x] = F[x \mid \theta]$).

Note that the basic probability measure P can be interpreted as P_{θ_0}, while the dependence of the distribution of the simulated statistics upon other values of the parameter θ is expressed by making $S_i(\theta)$ a function of θ (as well as $\omega \in \mathscr{X}$). In parametric models, the statistic S will usually be simulated by first generating an "observation" vector y according to an equation of the form

$$y = g(\theta, u), \tag{4.12}$$

where u has a known distribution (which can be simulated) and then computing

$$S(\theta) \equiv S[g(\theta, u)] \equiv g_S(\theta, u). \tag{4.13}$$

In such cases, the above assumptions can be interpreted as follows: $S_0 = S[y(\theta_0, u_0)]$ and $S_i(\theta) = S[y(\theta, u_i)]$, $i = 1, \ldots, N$, where the random vectors $u_0, u_1, \ldots, u_N$ are i.i.d. (or exchangeable). Note θ may include the parameters of a disturbance distribution in a model, such as covariance coefficients (or even its complete distribution function), so that the assumption that u has a known distribution is not restrictive. Assumptions on the structure of the parameter space Ω (e.g., whether it is finite-dimensional) will however entail real restrictions on the data-generating process. More generally, it is always possible to consider that the variables $S_0, S_1(\theta), \ldots, S_N(\theta)$ are P-measurable by considering their representation in terms of uniform random variables (see Shorack and Wellner, 1986, Chapter 1, Theorem 1): $S_0 = F^{-1}[V_0 \mid \theta_0]$ and $S_i(\theta) = F^{-1}[V_i \mid \theta]$, $i = 1, \ldots, N$, where $V_0, V_1, \ldots, V_N$ are P-exchangeable with uniform marginal distributions $[V_i \sim U(0, 1),\ i = 0, 1, \ldots, N]$.

A more general setup that allows for nonparametric models would consist in assuming that the null distribution of the test statistic depends on θ only through some transformation $T(y)$ of the observation vector y, which in turn only depends upon θ through some transformation $\theta_* = h(\theta)$, e.g. a subvector of θ:

$$T(y) = g[h(\theta), u] = g[\theta_*, u], \quad \theta_* \in \Omega_*, \tag{4.14}$$

where $\Omega_* = h(\Omega)$, hence

$$S(\theta) = S(T(y)) = S(g[h(\theta), u]) \equiv g_S[h(\theta), u] = g_S(\theta_*, u). \tag{4.15}$$

The setup (4.14)–(4.15) allows for reductions of the nuisance parameter space (e.g., through invariance). In particular, nonparametric models may be considered by taking appropriate distribution-free statistics (e.g., test statistics based on signs, ranks, permutations, etc.). What matters for our purpose is the possibility of simulating the test statistic, not necessarily the data themselves.

Let also

$$\hat{F}_N[x \mid \theta] \equiv \hat{F}_N[x; S(N, \theta)], \quad \hat{G}_N[x \mid \theta] \equiv \hat{G}_N[x; S(N, \theta)], \tag{4.16}$$

$$\hat{p}_N[x \mid \theta] = \frac{N\hat{G}_N[x \mid \theta] + 1}{N + 1} \tag{4.17}$$

be defined as in (2.8)–(2.10), and suppose the variables

$$\sup\{\hat{G}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \text{ and } \inf\{\hat{F}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \text{ are } \mathscr{A}_{\mathscr{X}}\text{-measurable where } \Omega_0 \text{ is nonempty subset of } \Omega. \tag{4.18}$$

For general discussions of measurability of conditions for extrema of random functions, the reader may consult Debreu (1967), Brown and Purves (1973) and Stinchcombe and White (1992).[3] We then get the following proposition.

Proposition 4.1 (*Validity of MMC tests when ties have zero probability*). *Under the assumptions and notations* (4.10), (4.11) *and* (4.16)–(4.18), *set* $S_0(\theta_0) = S_0$ *and suppose that*

$$\mathsf{P}[S_i(\theta_0) = S_j(\theta_0)] = 0 \quad \text{for } i \neq j, \ \ i, j = 0, 1, \ldots, N. \tag{4.19}$$

If $\theta_0 \in \Omega_0$, *then for* $0 \leqslant \alpha_1 \leqslant 1$,

$$\begin{aligned}\mathsf{P}[\sup\{\hat{G}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \leqslant \alpha_1] &\leqslant \mathsf{P}[\inf\{\hat{F}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \geqslant 1 - \alpha_1] \\ &\leqslant \frac{I[\alpha_1 N] + 1}{N + 1},\end{aligned} \tag{4.20}$$

where $\mathsf{P}[\inf\{\hat{F}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \geqslant 1 - \alpha_1] = \mathsf{P}[S_0 \geqslant \sup\{\hat{F}_N^{-1}[1 - \alpha_1 \mid \theta] : \theta \in \Omega_0\}]$ *for* $0 < \alpha_1 < 1$, *and*

$$\mathsf{P}[\sup\{\hat{p}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \leqslant \alpha] \leqslant \frac{I[\alpha(N + 1)]}{N + 1} \quad \text{for } 0 \leqslant \alpha \leqslant 1. \tag{4.21}$$

Following the latter proposition, if we choose α_1 and N so that (2.20) holds, the critical region $\sup\{\hat{G}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \leqslant \alpha_1$ has level α irrespective of the presence of nuisance parameters in the distribution of the test statistic S under the null hypothesis $H_0 : \theta_0 \in \Omega_0$. The same also holds if we use the (almost) equivalent critical regions $\inf\{\hat{F}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \geqslant 1 - \alpha_1$ or $S_0 \geqslant \sup\{\hat{F}_N^{-1}[1 - \alpha_1 \mid \theta] : \theta \in \Omega_0\}$. We shall call such tests MMC tests.

[3] If measurability is an issue, notions of "near-measurability" can be substituted (see Stinchcombe and White, 1992). From the viewpoint of getting upper bounds on probabilities, the probability operator can also be replaced by the associated outer measure which is always well-defined (see Dufour, 1989, Footnote 5).

To be more explicit, if $S(\theta)$ is generated according to expressions of the form (4.14)–(4.15), we have

$$\hat{G}_N[S_0 \mid \theta] = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[S_i(\theta) \geqslant S_0] = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[S(g[h(\theta), u_i]) \geqslant S_0]$$
$$= \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[g_S(\theta_*, u_i) \geqslant S_0]. \tag{4.22}$$

The function $\hat{G}_N[S_0 \mid \theta]$ (or $\hat{p}_N[S_0 \mid \theta]$), is then maximized with respect to $\theta \in \Omega_0$ [or equivalently, with respect to $\theta_* \in \Omega_{0*} = h(\Omega_0)$] keeping the observed statistic S_0 and the simulated disturbance vectors $u_1, \ldots, u_N$ fixed. The function $\hat{G}_N[S_0 \mid \theta]$ is a step-type function which typically has zero derivatives almost everywhere, except on isolated points (or manifolds) where it is not differentiable. Further, the supremum of $\hat{G}_N[S_0 \mid \theta]$ is typically not unique, in the sense that several values of θ will yield the required supremum. So it cannot be maximized with usual derivative-based algorithms. However, the required maximizations can be performed by using appropriate optimization algorithms that do not require differentiability, such as *simulated annealing*. For further discussion of such algorithms, the reader may consult Goffe et al. (1994).

It is easy to extend Proposition 4.1 in order to relax the no-tie assumption (4.19). For that purpose, we generate as in Proposition 2.4 a vector $(U_0, U_1, \ldots, U_N)$ of $N+1$ i.i.d. $U(0,1)$ random variables independent of $S_0, S_1(\theta_0), \ldots, S_N(\theta_0)$, and we consider the corresponding randomized distribution, tail area and p-value functions:

$$\tilde{F}_N[x \mid \theta] \equiv \tilde{F}_N[x; U_0, S(N,\theta), U(N)], \tag{4.23}$$

$$\tilde{G}_N[x \mid \theta] \equiv \tilde{G}_N[x; U_0, S(N,\theta), U(N)], \quad \tilde{p}_N[x \mid \theta] = \frac{N\tilde{G}_N[x \mid \theta] + 1}{N+1}, \tag{4.24}$$

where

$$U(N) = (U_1, \ldots, U_N) \quad \text{and} \quad U_0, U_1, \ldots, U_N \overset{\text{i.i.d.}}{\sim} U(0,1). \tag{4.25}$$

Under the corresponding measurability assumption

$$\sup\{\tilde{G}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \text{ and } \inf\{\tilde{F}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \text{ are } \mathscr{A}_{\mathscr{X}}\text{-measurable}$$
$$\text{where } \Omega_0 \text{ is nonempty subset of } \Omega, \tag{4.26}$$

we can then state the following generalization of Proposition 2.4.

Proposition 4.2 (*Validity of MMC tests for general statistics*). *Under the assumptions and notations* (4.10), (4.11), (4.16)–(4.18) *and* (4.23)–(4.26), *suppose* $U_0, U_1, \ldots, U_N$ *are independent of* $S_0, S_1(\theta_0), \ldots, S_N(\theta_0)$. *If* $\theta_0 \in \Omega_0$, *then for* $0 \leqslant \alpha_1 \leqslant 1$ *and for* $0 \leqslant \alpha \leqslant 1$,

$$\mathsf{P}[\sup\{\hat{G}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \leqslant \alpha_1] \leqslant \mathsf{P}[\sup\{\tilde{G}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \leqslant \alpha_1]$$
$$\leqslant \frac{I[\alpha_1 N] + 1}{N+1},$$

$$\mathsf{P}[\sup\{\hat{G}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \leqslant \alpha_1] \leqslant \mathsf{P}[\sup\{\tilde{F}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \geqslant 1 - \alpha_1] \leqslant \frac{I[\alpha_1 N] + 1}{N + 1},$$

$$\mathsf{P}[\sup\{\hat{p}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \leqslant \alpha] \leqslant \mathsf{P}[\sup\{\tilde{p}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \leqslant \alpha] \leqslant \frac{I[\alpha(N + 1)]}{N + 1}.$$

One should note here that the validity results of Propositions 4.1 and 4.2 differ from those of the corresponding Propositions 2.2 and 2.4 in the sense that equalities have been replaced by inequalities. This entails that the corresponding maximized MC tests is exact in the sense that the probability of type I error cannot be larger than the nominal level α of the test, but its size may be lower that the level (leading to a conservative procedure).[4] In view of the fact the distribution of the test statistic involves nuisance parameters, this is not surprising: since the distribution of the test statistic varies as a function of nuisance parameters, we can expect that the probability of type I error be lower than the level α for some distribution compatible with the null hypothesis, even if we use the tightest possible critical value that allows one to control the level of the test. Both the fundamental (infeasible) test and its MC version are *not similar*. This is a feature of the test statistic, not its MC implementation. Of course, it is preferable from the power viewpoint that the discrepancy between the size of the test and its level be as small as possible. This discrepancy would disappear if we could estimate and maximize without error the theoretical p-value function $G[x \mid \theta]$ or the appropriate critical value, but this is not typically feasible. In general, the discrepancy between the size and the nominal level of the test depends on the form of the test statistic, the null hypothesis, and the number N of replications of the MMC procedure. Studying in any detail this sort of effect would go beyond the scope of the present.[5]

5. Asymptotic Monte Carlo tests based on a consistent set estimator

In this section, we propose simplified approximate versions of the procedures proposed in the previous section when a consistent point or set estimate of θ is available. To do this, we shall need to reformulate the setup used previously in order to allow for an increasing sample size.

Consider

$$S_{T0}, S_{T1}(\theta), \ldots, S_{TN}(\theta), \quad T \geqslant I_0, \ \theta \in \Omega, \tag{5.1}$$

[4]We say that a test procedure is *conservative* at level α if its size is *strictly smaller* than α. Note that a non-similar test is not conservative as long as its size is equal to the level α (even though the probability of type I error is smaller than α for certain distributions compatible with the null hypothesis).

[5]A question of interest here consists in studying the conditions under which the discrepancy will disappear as the number of MC replications goes to infinity ($N \to \infty$). The reader will also find simulation evidence on the size and power properties of MMC procedures in Dufour and Khalaf (2003a, b), Dufour and Jouini (2005) and Dufour and Valéry (2005).

real random variables all defined on a common probability space $(\mathcal{X}, \mathcal{A}_{\mathcal{X}}, \mathsf{P})$, and set

$$S_T(N, \theta) = (S_{T1}(\theta), \ldots, S_{TN}(\theta)), \quad T \geqslant I_0. \tag{5.2}$$

We will be primarily interested by situations where

$$\text{the variables } S_{T0}, S_{T1}(\theta_0), \ldots, S_{TN}(\theta_0) \text{ are exchangeable for some } \theta_0 \in \Omega, \text{ each one with distribution function } F_T[x \mid \theta_0]. \tag{5.3}$$

Here S_{T0} will normally refer to a test statistic with distribution function $F_T[\cdot \mid \theta]$ based on a sample of size T, while $S_{T1}(\theta), \ldots, S_{TN}(\theta)$ are i.i.d. replications of the same test statistic obtained independently under the assumption that the parameter vector is θ (i.e., $\mathsf{P}[S_{Ti}(\theta) \leqslant x] = F_T[x \mid \theta], \ i = 1, \ldots, N$). Let also

$$\hat{F}_{TN}[x \mid \theta] = \hat{F}_N[x; S_T(N, \theta)], \quad \hat{G}_{TN}[x \mid \theta] = \hat{G}_N[x; S_T(N, \theta)], \tag{5.4}$$

$$\hat{p}_{TN}[x \mid \theta] = \frac{N\hat{G}_{TN}[x \mid \theta] + 1}{N + 1}, \tag{5.5}$$

and let $\hat{F}_{TN}^{-1}[x \mid \theta]$ be defined as in (2.7)–(2.10).

We consider first the situation where p-values are maximized over a subset C_T of Ω (e.g., a confidence set for θ) instead of Ω_0. Consequently, we introduce the following assumption:

$$\begin{aligned} &C_T, T \geqslant I_0 \text{ is a sequence of (possibly random) subsets of } \Omega \text{ such that} \\ &\quad \sup\{\hat{G}_{TN}[S_{T0} \mid \theta] : \theta \in C_T\} \text{ and} \\ &\quad \inf\{\hat{F}_{TN}[S_{T0} \mid \theta] : \theta \in C_T\} \text{ are } \mathcal{A}_{\mathcal{X}}\text{-measurable,} \\ &\qquad \text{for all } T \geqslant I_0, \text{ where } \Omega_0 \text{ is nonempty subset of } \Omega. \end{aligned} \tag{5.6}$$

Then we have the following proposition.

Proposition 5.1 (*Asymptotic validity of confidence-set restricted MMC tests: continuous distributions*). *Under the assumptions and notations* (5.1) *to* (5.6), *set* $S_{T0}(\theta_0) = S_{T0}$, *suppose*

$$\mathsf{P}[S_{Ti}(\theta_0) = S_{Tj}(\theta_0)] = 0 \quad \textit{for } i \neq j \quad \textit{and} \quad i, j = 0, 1, \ldots, N, \tag{5.7}$$

for all $T \geqslant I_0$, *and let* $C_T, T \geqslant I_0$, *be a sequence of* (*possibly random*) *subsets of* Ω *such that*

$$\lim_{T \to \infty} \mathsf{P}[\theta_0 \in C_T] = 1. \tag{5.8}$$

If $\theta_0 \in \Omega_0$, *then*

$$\begin{aligned} &\lim_{T \to \infty} \mathsf{P}[\sup\{\hat{G}_{TN}[S_{T0} \mid \theta] : \theta \in C_T\} \leqslant \alpha_1] \\ &\quad \leqslant \lim_{T \to \infty} \mathsf{P}[\inf\{\hat{F}_{TN}[S_{T0} \mid \theta] : \theta \in C_T\} \geqslant 1 - \alpha_1] \\ &\quad = \lim_{T \to \infty} \mathsf{P}[S_{T0} \geqslant \sup\{\hat{F}_{TN}^{-1}[1 - \alpha_1 \mid \theta] : \theta \in C_T\}] \leqslant \frac{I[\alpha_1 N] + 1}{N + 1} \end{aligned} \tag{5.9}$$

and

$$\lim_{T\to\infty} \mathsf{P}[\sup\{\hat{p}_{TN}[S_{T0} \mid \theta] : \theta \in C_T\} \leqslant \alpha] \leqslant \frac{I[\alpha(N+1)]}{N+1} \quad \textit{for } 0 \leqslant \alpha \leqslant 1. \tag{5.10}$$

It is quite easy to find a consistent set estimate of θ_0 whenever a consistent point estimate $\hat{\theta}_T$ of θ_0 is available. Suppose $\Omega \subseteq \mathbb{R}^k$ and

$$\lim_{T\to\infty} \mathsf{P}[\|\hat{\theta}_T - \theta_0\| < \varepsilon] = 1, \quad \forall \varepsilon > 0, \tag{5.11}$$

where $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^k$ [i.e., $\|x\| = (x'x)^{1/2}$, $x \in \mathbb{R}^k$]. Note that condition (5.8) need only hold for the true value θ_0 of the parameter vector θ. Then any set of the form $C_T = \{\theta \in \Omega : \|\hat{\theta}_T - \theta\| < c\}$ satisfies (5.8), whenever c is a fixed positive constant that does not depend on T. More generally, if there is a sequence of (possibly random) matrices A_T and a non-negative exponent δ such that

$$\lim_{T\to\infty} \mathsf{P}[T^{\delta}\|A_T(\hat{\theta}_T - \theta_0)\|^2 < c] = 1, \quad \forall c > 0, \tag{5.12}$$

then any set of the form

$$\begin{aligned} C_T &= \{\theta \in \Omega : (\hat{\theta}_T - \theta)' A_T' A_T (\hat{\theta}_T - \theta) < c/T^{\delta}\} \\ &= \{\theta \in \Omega : \|A_T(\hat{\theta}_T - \theta)\|^2 < c/T^{\delta}\}, \quad c > 0 \end{aligned} \tag{5.13}$$

satisfies (5.8), since in this case,

$$\begin{aligned} \mathsf{P}[\theta_0 \in C_T] &= \mathsf{P}[(\hat{\theta}_T - \theta_0)' A_T' A_T (\hat{\theta}_T - \theta_0) < c/T^{\delta}] \\ &= \mathsf{P}[T^{\delta}(\hat{\theta}_T - \theta_0)' A_T' A_T (\hat{\theta}_T - \theta_0) < c] \underset{T\to\infty}{\longrightarrow} 1. \end{aligned}$$

In particular (5.12) will hold whenever we can find $\bar{\delta} > 0$ (e.g., $\bar{\delta} = 1$) such that $T^{\bar{\delta}/2} A_T(\hat{\theta}_T - \theta_0)$ has an asymptotic distribution (as $T \to \infty$) and δ is selected so that $0 \leqslant \delta < \bar{\delta}$. Whenever $\delta > 0$ and $plim_{T\to\infty}(A_T' A_T) = C_0$ with $\det(C_0) \neq 0$, the diameter of the set C_T goes to zero, a fact which can greatly simplify the evaluation of the variables $\sup\{\hat{G}_{TN}\}$, $\inf\{\hat{F}_{TN}\}$ and $\sup\{\hat{p}_{TN}\}$ in (5.9) and (5.10).

The above procedure may be especially useful when the distribution of the test statistic is highly sensitive to nuisance parameters, in a way that would make its asymptotic distribution discontinuous with respect to the nuisance parameters. In such cases, a simulation-based procedure where the nuisance parameters are replaced by a consistent point estimate—such as a parametric bootstrap procedure—may not converge to the appropriate asymptotic distribution (because the point estimate does not converge sufficiently fast to overcome the discontinuity). Here, possible discontinuities in the asymptotic distribution are automatically taken into account thorough a numerical maximization over a set that contains the correct value of the nuisance parameter with a probability asymptotically equal to one: using a consistent set estimator as opposed a point estimate (which does not converge fast enough) can overcome such a high sensitivity to nuisance parameters. Of course, the procedure can also be helpful in situations where the finite-sample distribution is highly sensitive to nuisance parameters, even though it does not lead to asymptotic failure of the bootstrap.

Again, it is possible to extend Proposition 5.1 to statistics with general (possibly discrete) distributions by considering properly randomized distribution, tail-area and p-value functions:

$$\tilde{F}_{TN}[x \mid \theta] = \tilde{F}_N[x; U_0, S_T(N;\theta), U(N)], \tag{5.14}$$

$$\tilde{G}_{TN}[x \mid \theta] = \tilde{G}_N[x; U_0, S_T(N;\theta), U(N)], \tag{5.15}$$

$$\tilde{p}_{TN}[x \mid \theta] = \frac{N\tilde{G}_{TN}[x \mid \theta] + 1}{N+1}, \tag{5.16}$$

where $\tilde{F}_N[\cdot]$, $\tilde{G}_N[\cdot]$, U_0 and $U(N)$ are defined as in (4.23)–(4.25).

Proposition 5.2 (*Asymptotic validity of confidence-set restricted MMC tests: general distributions*). *Under the assumptions and notations* (5.1)–(5.6) *and* (5.14)–(5.16), *suppose the sets* $C_T \subseteq \Omega$, $T \geqslant I_0$, *satisfy* (5.8). *If* $\theta_0 \in \Omega_0$, *then for* $0 \leqslant \alpha_1 \leqslant 1$ *and* $0 \leqslant \alpha \leqslant 1$,

$$\begin{aligned} &\lim_{T\to\infty} \mathsf{P}[\sup\{\hat{G}_{TN}[S_{T0} \mid \theta] : \theta \in C_T\} \leqslant \alpha_1] \\ &\quad \leqslant \lim_{T\to\infty} \mathsf{P}[\sup\{\tilde{G}_{TN}[S_{T0} \mid \theta] : \theta \in C_T\} \leqslant \alpha_1] \leqslant \frac{I[\alpha_1 N] + 1}{N+1}, \end{aligned} \tag{5.17}$$

$$\begin{aligned} &\lim_{T\to\infty} \mathsf{P}[\sup\{\hat{G}_{TN}[S_{T0} \mid \theta] : \theta \in C_T\} \leqslant \alpha_1] \\ &\quad \leqslant \lim_{T\to\infty} \mathsf{P}[\sup\{\tilde{F}_{TN}[S_{T0} \mid \theta] : \theta \in C_T\} \geqslant 1 - \alpha_1] \leqslant \frac{I[\alpha_1 N] + 1}{N+1}, \end{aligned} \tag{5.18}$$

$$\begin{aligned} \lim_{T\to\infty} \mathsf{P}[\sup\{\hat{p}_{TN}[S_{T0} \mid \theta] : \theta \in C_T\} \leqslant \alpha] &\leqslant \lim_{T\to\infty} \mathsf{P}[\sup\{\tilde{p}_{TN}[S_0 \mid \theta] : \theta \in C_T\} \leqslant \alpha] \\ &\leqslant \frac{I[\alpha(N+1)]}{N+1}. \end{aligned} \tag{5.19}$$

6. Asymptotic Monte Carlo tests based on consistent point estimate

Parametric bootstrap tests may be interpreted as a simplified form of the procedures described in Propositions 5.1 and 5.2 where the consistent confidence set C_T has been replaced by a consistent point estimate $\hat{\theta}_T$. In other words, the distribution of $S_T(\theta)$, $\theta \in \Omega_0$, is simulated at a single point $\hat{\theta}_T$, leading to a *local* (or *pointwise*) MC test. It is well known that such bootstrap tests are not generally valid, unless stronger regularity conditions are imposed. In the following proposition, we extend earlier proofs of the asymptotic validity of such bootstrap tests. In particular, we allow for the presence of nuisance parameters in the asymptotic distribution of the test statistic considered. Further, our proofs have the interesting feature of being cast in the MC test setup where the number of replications N is kept fixed even asymptotically.

Such pointwise procedures require stronger regularity assumptions (such as uniform continuity and convergence over the nuisance parameter space)—so that they may fail in irregular cases where the maximized procedures described in the previous sections succeed in controlling the level of the test (at least asymptotically). But they are simpler to implement and may be taken as a natural starting point for implementing maximized procedures. In particular, if a pointwise MC p-value is *larger* than the level α of the test (so that the *pointwise MC test is not significant* at level α), it is clear that the *maximized* p-value must be larger than α (so that the *maximized MC test is not significant* at level α).

In order to establish a clear asymptotic validity result, we will use four basic assumptions on the distributions of the statistics $S_T(\theta)$ as functions of the parameter vector θ:

$$S_{T1}(\theta), \ldots, S_{TN}(\theta) \text{ are i.i.d. according to the distribution} \\ F_T[x \mid \theta] = \mathsf{P}[S_T(\theta) \leqslant x], \quad \forall \theta \in \Omega, \tag{6.1}$$

$$\Omega \text{ is a nonempty subset of } \mathbb{R}^k, \tag{6.2}$$

$$\text{for every} T \geqslant I_0,\ S_{T0} \text{ is a real random variable and } \hat{\theta}_T \text{ an estimator of } \theta, \\ \text{both measurable with respect to the probability space } (\mathscr{X}, \mathscr{A}_{\mathscr{X}}, \mathsf{P}), \\ \text{and } F_T[S_{T0} \mid \hat{\theta}_T] \text{ is a random variable;} \tag{6.3}$$

$$\forall \varepsilon_0 > 0,\ \forall \varepsilon_1 > 0,\ \exists \delta > 0 \text{ and a sequence of open subsets } D_{T0}(\varepsilon_0) \text{ in } \mathbb{R} \text{ such that} \\ \liminf_{T \to \infty} \mathsf{P}[S_{T0} \in D_{T0}(\varepsilon_0)] \geqslant 1 - \varepsilon_0 \text{ and} \\ \|\theta - \theta_0\| \leqslant \delta \Rightarrow \limsup_{T \to \infty} \left\{ \sup_{x \in D_{T0}(\varepsilon_0)} |F_T[x \mid \theta] - F_T[x \mid \theta_0]| \right\} \leqslant \varepsilon_1. \tag{6.4}$$

The first of these four conditions replaces the exchangeability assumption by an assumption of i.i.d. variables. The two next ones simply make appropriate measurability assumptions, while the last one may be interpreted as a local equicontinuity condition (at $\theta = \theta_0$) on the sequence of distribution functions $F_T(x \mid \theta)$, $T \geqslant I_0$. Note that S_{T0} is not assumed to follow the same distribution as the other variables $S_{T1}(\theta), \ldots, S_{TN}(\theta)$. Furthermore, S_{T0} and $F_T(x \mid \theta)$ do not necessarily converge to limits as $T \to \infty$. An alternative assumption of interest would consist in assuming that S_{T0} converges in probability ($\xrightarrow{p}$) to a random variable $\bar{S}_0$ as $T \to \infty$, in which case the "global" equicontinuity condition (6.4) can be weakened to a "local" one:

$$S_{T0} \xrightarrow[T \to \infty]{p} \bar{S}_0, \tag{6.5}$$

$$D_0 \text{ is a subset of } \mathbb{R} \text{ such that } \mathsf{P}[\bar{S}_0 \in D_0 \text{ and } S_{T0} \in D_0 \text{ for all } T \geqslant I_0] = 1, \tag{6.6}$$

$\forall x \in D_0$, $\forall \varepsilon > 0$, $\exists \delta > 0$ and an open neighborhood $B(x, \varepsilon)$ of x such that

$$\|\theta - \theta_0\| \leqslant \delta \Rightarrow \limsup_{T\to\infty} \left\{ \sup_{y \in B(x,\varepsilon)\cap D_0} |F_T[y \mid \theta] - F_T[y \mid \theta_0]| \right\} \leqslant \varepsilon. \tag{6.7}$$

We can now show that Monte Carlo tests obtained by simulating $S_{Ti}(\theta)$, $i = 1, \ldots, N$, with $\theta = \hat{\theta}_T$ are equivalent for large T to those based on using the true value $\theta = \theta_0$.

Proposition 6.1 (*Asymptotic validity of bootstrap p-values*). *Under the assumptions and notations* (5.1), (5.2), (5.4), (5.5), (5.14)–(5.16) *and* (6.1)–(6.3), *suppose the random variable S_{T0} and the estimator $\hat{\theta}_T$ are both independent of $S_T(N, \theta)$ and U_0. If* $\operatorname{plim}_{T\to\infty} \hat{\theta}_T = \theta_0$ *and condition* (6.4) *or* (6.5)–(6.7) *hold, then for $0 \leqslant \alpha_1 \leqslant 1$ and $0 \leqslant \alpha \leqslant 1$,*

$$\begin{aligned} &\lim_{T\to\infty} \{\mathsf{P}[\tilde{G}_{TN}[S_{T0} \mid \hat{\theta}_T] \leqslant \alpha_1] - \mathsf{P}[\tilde{G}_{TN}[S_{T0} \mid \theta_0] \leqslant \alpha_1]\} \\ &\quad = \lim_{T\to\infty} \{\mathsf{P}[\hat{G}_{TN}[S_{T0} \mid \hat{\theta}_T] \leqslant \alpha_1] - \mathsf{P}[\hat{G}_{TN}[S_{T0} \mid \theta_0] \leqslant \alpha_1]\} = 0 \end{aligned} \tag{6.8}$$

and

$$\begin{aligned} &\lim_{T\to\infty} \{\mathsf{P}[\tilde{p}_{TN}[S_{T0} \mid \hat{\theta}_T] \leqslant \alpha] - \mathsf{P}[\tilde{p}_{TN}[S_{T0} \mid \theta_0] \leqslant \alpha]\} \\ &\quad = \lim_{T\to\infty} \{\mathsf{P}[\hat{p}_{TN}[S_{T0} \mid \hat{\theta}_T] \leqslant \alpha] - \mathsf{P}[\hat{p}_{TN}[S_{T0} \mid \theta_0] \leqslant \alpha]\} = 0. \end{aligned} \tag{6.9}$$

It is worth noting that condition (6.7) holds whenever $F_T[x \mid \theta]$ converges to a distribution function $F_\infty[x \mid \theta]$ which is continuous with respect to $(x, \theta')'$, for $x \in D_0$, as follows:

$\forall x \in D_0$, $\forall \varepsilon > 0$, $\exists \delta_1 > 0$ and an open neighborhood $B_1(x, \varepsilon)$ of x such that

$$\|\theta - \theta_0\| \leqslant \delta \Rightarrow \limsup_{T\to\infty} \left(\sup_{y \in B_1(x,\varepsilon)\cap D_0} |F_T[y \mid \theta] - F_\infty[y \mid \theta]| \right) \leqslant \varepsilon, \tag{6.10}$$

$\forall x \in D_0$, $\forall \varepsilon > 0$, $\exists \delta_2 > 0$ and an open neighborhood $B_2(x, \varepsilon)$ of x such that

$$\|\theta - \theta_0\| \leqslant \delta_2 \Rightarrow \sup_{y \in B_2(x,\varepsilon)\cap D_0} |F_\infty[y \mid \theta] - F_\infty[y \mid \theta_0]| \leqslant \varepsilon. \tag{6.11}$$

It is then easy to see that (6.10)–(6.11) entail (6.7) on noting that

$$\begin{aligned} |F_T[x \mid \theta] - F_T[x \mid \theta_0]| \leqslant |F_T[x \mid \theta] - F_\infty[x \mid \theta]| + |F_T[x \mid \theta_0] - F_\infty[x \mid \theta_0]| \\ + |F_\infty[x \mid \theta] - F_\infty[x \mid \theta_0]|, \quad \forall x. \end{aligned}$$

Note also that (6.11) holds whenever $F_\infty[x \mid \theta]$ is continuous with respect to $(x, \theta')'$, although the latter condition is not at all necessary (e.g., in models where D_0 is a discrete set of points). In particular, (6.10)–(6.11) will hold when $F_T[x \mid \theta]$ admits an expansion around a pivotal distribution:

$$F_T[x \mid \theta] = F_\infty(x) + T^{-\gamma} g(x, \theta) + h_T(x, \theta), \tag{6.12}$$

where $F_\infty(x)$ is a distribution function that does not depend on $\theta, \gamma > 0$, with the following assumptions on $g(x, \theta)$ and $h_T(x, \theta)$:

$$
\begin{aligned}
&\forall x \in D_0,\ \exists \text{ an open neighborhood } B(x, \theta_0) \text{ of } (x, \theta_0')' \text{ such that} \\
&\quad |g(y, \theta)| \leqslant C(x, \theta), \quad \text{for all } (y, \theta')' \in B(x, \theta_0) \cap D_0, \\
&\qquad \text{where } C(x, \theta) \text{ is a positive constant, and} \\
&\quad T^\gamma h_T(y, \theta) \underset{T\to\infty}{\longrightarrow} 0 \text{ uniformly on } B(x, \theta_0) \cap D_0.
\end{aligned} \tag{6.13}
$$

The latter is quite similar (although somewhat weaker) to the assumption considered by Hall and Titterington (1989, Eq. (2.5)).

When S_{T0} is distributed like $S_T(\theta_0)$, i.e., $\mathsf{P}[S_{T0} \leqslant x] = F_T[x \mid \theta_0]$, we can apply Proposition 2.4 and see that $\mathsf{P}[\tilde{G}_{TN}[S_{T0} \mid \theta_0] \leqslant \alpha_1] = (I[\alpha_1 N] + 1)/(N + 1)$, hence

$$
\lim_{T\to\infty} \mathsf{P}[\tilde{G}_{TN}[S_{T0} \mid \hat{\theta}_T] \leqslant \alpha_1] = \frac{I[\alpha_1 N] + 1}{N + 1}. \tag{6.14}
$$

7. Conclusion

In this paper, we have made four main contributions. First, for the case where we have a test statistic whose distribution does not involve nuisance parameters under the null hypothesis, we have proposed a general form of Monte Carlo testing which allows for exchangeable (as opposed to i.i.d.) Monte Carlo replications of general test statistics whose distribution can take an arbitrary form (continuous, discrete or mixed). In particular, this form is not limited to permutation tests, which has received considerable attention in the earlier literature on Monte Carlo tests (see Dwass, 1957; Green, 1977; Vadiveloo, 1983; Keller-McNulty and Higgins, 1987; Lock, 1991; Edgington, 1995; Manly, 1997; Noreen, 1989; Good, 1994]. Second, we have shown how the method can be extended to models with nuisance parameters as long as the null distribution of the test statistic can be simulated once the nuisance parameters have been specified. This leads to what we called *maximized Monte Carlo tests* which were shown to satisfy the level constraint. Thirdly, we proposed a simplified version of the latter method which can lead to asymptotically valid tests, even if the asymptotic distribution depends on nuisance parameters in a discontinuous way. This method only requires one to use a consistent set estimator of the nuisance parameters, which is always feasible as long as a consistent point estimate of the nuisance parameters is available. Further, in the latter case, no additional information is required on the asymptotic distribution of the consistent estimator. Fourth, we showed that Monte Carlo tests obtained upon replacing unknown nuisance parameters by consistent estimates also lead to asymptotically valid tests. However, it is important to note that stronger conditions are needed for this to occur and such conditions may be difficult to check in practice.

The main shortcoming of the proposed MMC tests comes from the fact that such tests may be computationally demanding. We cannot study here the appropriate numerical algorithms or detailed implementations of the theory described above. But a number of such applications are presented in companion papers (which are based on earlier versions of the present paper). For example, for an illustration of the adjustment for discreteness proposed here, the reader may consult Dufour et al. (1998) where it is used to correct the size of Kolmogorov–Smirnov tests (which involve a discrete statistic) for the normality of errors in a linear regression. The method of maximized Monte Carlo tests can of course be applied to a wide array of models where nuisance parameters problems show up: for example, inference in seemingly unrelated regressions (Dufour and Khalaf, 2003a), simultaneous equations models (Dufour and Khalaf, 2003b), dynamic models (Dufour and Jouini, 2005; Dufour and Valéry, 2005), and models with limited dependent variables (Jouneau-Sion and Torrès, 2005). It is clear many more applications are possible. The size and power properties of the proposed procedures are also studied by simulation methods in this work.

Acknowledgements

The author thanks Torben Andersen, Benoît Perron, Russell Davidson, V.P. Godambe, Joel Horowitz, Joanna Jasiak, Linda Khalaf, Jan Kiviet, James MacKinnon, Charles Manski, Eric Renault, Christopher Sims, Olivier Torrès, Pascale Valéry, Michael Veall, Mark Watson, and two anonymous referees for several useful comments. This work was supported by the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Alexander-von-Humboldt Foundation (Germany), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Canada Council for the Arts (Killam Fellowship), the Natural Sciences and Engineering Research Council of Canada, the Social Scienes and Humanities Research Council of Canada, the Fonds de recherche sur la société et la culture (Québec), and the Fonds de recherche sur la nature et les technologies (Québec).

Appendix. Proofs

Proof of Lemma 2.1. By condition (2.12), the variables $y_1, y_2, \ldots, y_N$ are all distinct with probability 1, and the rank vector $(R_1, R_2, \ldots, R_N)'$ is with probability 1 a permutation of the integers $(1, 2, \ldots, N)$. Furthermore, since the variables $y_1, y_2, \ldots, y_N$ are exchangeable, the $N!$ distinct permutations of $(y_1, y_2, \ldots, y_N)$ have the same probability $1/N!$. Consequently, we have:

$$\mathsf{P}(R_j = i) = \frac{1}{N}, \quad i = 1, 2, \ldots, N,$$
$$\mathsf{P}(R_j/N \leqslant x) = I[xN]/N, \quad 0 \leqslant x \leqslant 1,$$

from which (2.13) follows and

$$\begin{aligned} \mathsf{P}(R_j/N < x) &= (I[xN]-1)/N \quad \text{if } xN \in \mathbb{Z}_+, \\ &= I[xN]/N \quad \text{otherwise,} \end{aligned}$$

where $\mathbb{Z}_+$ is the set of the positive integers. Since, for any real number z,

$$\begin{aligned} I[N-z] &= N-z \quad \text{if } z \text{ is an integer,} \\ &= N - I[z] - 1 \quad \text{otherwise,} \end{aligned}$$

we then have, for $0 \leqslant x \leqslant 1$,

$$\begin{aligned} \mathsf{P}(R_j/N \geqslant x) &= 1 - \mathsf{P}(R_N/N < x) = (N - I[xN] + 1)/N \quad \text{if } xN \in \mathbb{Z}_+, \\ &= (N - I[xN])/N \quad \text{otherwise,} \end{aligned}$$

hence

$$\begin{aligned} \mathsf{P}[R_j/N \geqslant x] &= (I[(1-x)N]+1)/N \quad \text{if } 0 < x \leqslant 1, \\ &= 1 \quad \text{if } x = 0, \end{aligned}$$

from which we get (2.14). $\square$

Proof of Proposition 2.2. Assuming there are no ties among $S_0, S_1, \ldots, S_N$ (an event with probability 1), we have

$$\begin{aligned} \hat{G}_N(S_0) &= \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(S_i \geqslant S_0) = \frac{1}{N}\sum_{i=1}^{N}[1 - \mathbf{1}(S_i \leqslant S_0)] = 1 - \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(S_i \leqslant S_0) \\ &= 1 - \frac{1}{N}\left[-1 + \sum_{i=0}^{N} \mathbf{1}(S_i \leqslant S_0)\right] = (N+1-R_0)/N, \end{aligned}$$

where $R_0 = \sum_{i=0}^{N} \mathbf{1}(S_i \leqslant S_0)$ is the rank of S_0 when the $N+1$ variables $S_0, S_1, \ldots, S_N$ are ranked in nondecreasing order. Using (2.15) and Lemma 2.1, it then follows that

$$\begin{aligned} \mathsf{P}[\hat{G}_N(S_0) \leqslant \alpha_1] &= \mathsf{P}\left[\frac{N+1-R_0}{N} \leqslant \alpha_1\right] = \mathsf{P}\left[\frac{R_0}{N+1} \geqslant \frac{(1-\alpha_1)N+1}{N+1}\right] \\ &= \frac{I[\alpha_1 N]+1}{N+1} \end{aligned}$$

for $0 \leqslant \alpha_1 \leqslant 1$. Furthermore, $\hat{F}_N(S_0) = 1 - \hat{G}_N(S_0)$ with probability 1, and (2.17) follows. We then get (2.18) on observing that: $\hat{F}_N(y) \geqslant q \Longleftrightarrow y \geqslant \hat{F}_N^{-1}(q)$ for $y \in \mathbb{R}$ and $0 < q < 1$ (see Reiss, 1989, Appendix 1). Finally, to obtain (2.19), we note that

$$\hat{p}_N(S_0) \equiv \frac{N\hat{G}_N(S_0)+1}{N+1} \leqslant \alpha \Longleftrightarrow \hat{G}_N(S_0) \leqslant \frac{\alpha(N+1)-1}{N}$$

hence, since that $0 \leqslant \hat{G}_N(S_0) \leqslant 1$ and using (2.17),

$$\mathsf{P}[\hat{p}_N(S_0) \leqslant \alpha] = \mathsf{P}\left[\hat{G}_N(S_0) \leqslant \frac{\alpha(N+1)-1}{N}\right]$$
$$= \begin{cases} 0, & \text{if } \alpha < 1/(N+1), \\ \dfrac{I[\alpha(N+1)-1]+1}{N+1} = \dfrac{I[\alpha(N+1)]}{N+1}, & \text{if } \dfrac{1}{N+1} \leqslant \alpha \leqslant 1, \\ 1, & \text{if } \alpha > 1, \end{cases}$$

from which (2.19) follows upon observing that $I[\alpha(N+1)] = 0$ for $0 \leqslant \alpha < 1/(N+1)$. □

Proof of Lemma 2.3. From (2.23) and the continuity of the $U(0,1)$ distribution, we see easily that $\mathsf{P}[(y_i, U_i) = (y_j, U_j)] \leqslant \mathsf{P}[U_i = U_j] = 0$, for $i \neq j$, from which it follows that $\mathsf{P}[\tilde{R}_i = \tilde{R}_j] = 0$ for $i \neq j$ and the rank vector $(\tilde{R}_1, \tilde{R}_2, \ldots, \tilde{R}_N)$ is with probability 1 a random permutation of the integers $(1, 2, \ldots, N)$. Set $V_i = (y_i, U_i),\ i = 1, \ldots, N$. By considering all possible permutations $(V_{r_1}, V_{r_2}, \ldots, V_{r_N})$ of $(V_1, V_2, \ldots, V_N)$, and since $(V_{r_1}, V_{r_2}, \ldots, V_{r_N}) \sim (V_1, V_2, \ldots, V_N)$ for all permutations $(r_1, r_2, \ldots, r_N)$ of $(1, 2, \ldots, N)$ [by the exchangeability assumption], the elements of $(\tilde{R}_1, \tilde{R}_2, \ldots, \tilde{R}_N)$ are also exchangeable.The result then follows from Lemma 2.1. The reader may note that an alternative proof could be obtained by modifying the proof of Theorem 29A of Hájek (1969) to relax the independence assumption for $y_1, \ldots, y_N$. □

Proof of Proposition 2.4. Since the pairs $(S_i, U_i),\ i = 0, 1, \ldots, N$, are all distinct with probability 1, we have almost surely:

$$\tilde{G}_N(S_0) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[(S_i, U_i) \geqslant (S_0, U_0)] = 1 - \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[(S_i, U_i) \leqslant (S_0, U_0)]$$
$$= 1 - \frac{1}{N}\left\{-1 + \sum_{i=0}^{N} \mathbf{1}[(S_i, U_i) \leqslant (S_0, U_0)]\right\} = (N+1-\tilde{R}_0)/N,$$

where $\tilde{R}_0 = \sum_{i=0}^{N} \mathbf{1}[(S_i, U_i) \leqslant (S_0, U_0)]$ is the randomized rank of S_0 obtained when ranking in ascending order [according to (2.23)] the $N+1$ pairs $(S_i, U_i),\ i = 0, 1, \ldots, N$. Using Lemma 2.3, it follows that

$$\mathsf{P}[\tilde{G}_N(S_0) \leqslant \alpha_1] = \mathsf{P}[(N+1-\tilde{R}_0)/N \leqslant \alpha_1] = \mathsf{P}\left[\frac{\tilde{R}_0}{N+1} \geqslant \frac{(1-\alpha_1)N+1}{N+1}\right]$$
$$= \begin{cases} 0 & \text{if } \alpha_1 < 0, \\ \dfrac{I[\alpha_1 N]+1}{N+1} & \text{if } 0 \leqslant \alpha_1 \leqslant 1, \\ 1 & \text{if } \alpha_1 > 1. \end{cases}$$

Since the pairs $(S_i, U_i),\ i = 0, 1, \ldots, N$, are all distinct with probability 1, we also have $\tilde{F}_N(S_0) = 1 - \tilde{G}_N(S_0)$ with probability 1, hence using inequalities (2.33)–(2.36),

$$\mathsf{P}[\hat{G}_N(S_0) \leqslant \alpha_1] \leqslant \mathsf{P}[\tilde{G}_N(S_0) \leqslant \alpha_1] = \mathsf{P}[\tilde{F}_N(S_0) \geqslant 1 - \alpha_1] \leqslant \mathsf{P}[\hat{F}_N(S_0) \geqslant 1 - \alpha_1]$$

and (2.36) is established. The identity $\mathsf{P}[\hat{F}_N(S_0) \geqslant 1-\alpha_1] = \mathsf{P}[S_0 \geqslant \hat{F}_N^{-1}(1-\alpha_1)]$ follows from the equivalence: $\hat{F}_N(y) \geqslant q \Longleftrightarrow y \geqslant \hat{F}_N^{-1}(q)$, $\forall y \in \mathbb{R}$, $0<q<1$. Finally, to obtain (2.37), we observe that

$$\tilde{p}_N(S_0) = \frac{N\tilde{G}_N(S_0)+1}{N+1} \leqslant \alpha \Longleftrightarrow \tilde{G}_N(S_0) \leqslant \frac{\alpha(N+1)-1}{N}$$

hence, using (2.36),

$$\begin{aligned}\mathsf{P}[\tilde{p}_N(S_0) \leqslant \alpha] &= \mathsf{P}\left[\tilde{G}_N(S_0) \leqslant \frac{\alpha(N+1)-1}{N}\right] \\ &= \begin{cases} 0 & \text{if } \alpha < 1/(N+1), \\ \dfrac{I[\alpha(N+1)-1]+1}{N+1} = \dfrac{I[\alpha(N+1)]}{N+1} & \text{if } \dfrac{1}{N+1} \leqslant \alpha \leqslant 1, \end{cases}\end{aligned}$$

from which (2.19) follows on observing that $I[\alpha(N+1)] = 0$ for $0 \leqslant \alpha \leqslant 1/(N+1)$. □

Proof of Proposition 4.1. Since

$$\hat{G}_N[S_0 \mid \theta] \geqslant 1 - \hat{F}_N[S_0 \mid \theta], \tag{A.1}$$

we have

$$\mathsf{P}[\sup\{\hat{G}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \leqslant \alpha_1] \leqslant \mathsf{P}[\inf\{\hat{F}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \geqslant 1-\alpha_1].$$

When $\theta_0 \in \Omega_0$, it is also clear that: $\inf_{\theta\in\Omega_0} \hat{F}_N[S_0 \mid \theta] \geqslant 1-\alpha_1 \Rightarrow \hat{F}_N[S_0 \mid \theta_0] \geqslant 1-\alpha_1$. Thus, using Proposition 2.2,

$$\mathsf{P}[\inf\{\hat{F}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \geqslant 1-\alpha_1] \leqslant \mathsf{P}[\hat{F}_N[S_0 \mid \theta_0] \geqslant 1-\alpha_1] = \frac{I[\alpha_1 N]+1}{N+1}.$$

Furthermore,

$$\begin{aligned} &\inf_{\theta\in\Omega_0} \hat{F}_N[S_0 \mid \theta] \geqslant 1-\alpha_1 \Longleftrightarrow \hat{F}_N[S_0 \mid \theta] \geqslant 1-\alpha_1, \quad \forall\theta \in \Omega_0 \\ &\Longleftrightarrow S_0 \geqslant \hat{F}_N^{-1}[1-\alpha_1 \mid \theta], \quad \forall\theta \in \Omega_0 \Longleftrightarrow S_0 \geqslant \sup_{\theta\in\Omega_0} \hat{F}_N^{-1}[1-\alpha_1 \mid \theta] \end{aligned} \tag{A.2}$$

so that, using Proposition 2.2,

$$\begin{aligned}\mathsf{P}[S_0 \geqslant \sup\{\hat{F}_N^{-1}[1-\alpha_1 \mid \theta_0] : \theta \in \Omega_0\}] &= \mathsf{P}[\inf\{\hat{F}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \geqslant 1-\alpha_1] \\ &\leqslant \frac{I[\alpha_1 N]+1}{N+1}\end{aligned}$$

and (4.20) is established. Eq. (4.21) follows in the same way on observing that $\sup_{\theta \in \Omega_0} \tilde{p}_N[S_0 \mid \theta] \leqslant \sup_{\theta \in \Omega_0} \hat{p}_N[S_0 \mid \theta]$ and

$$\sup_{\theta \in \Omega_0} \tilde{p}_N[S_0 \mid \theta] \leqslant \alpha \Rightarrow \tilde{p}_N[S_0 \mid \theta_0] \leqslant \alpha, \quad \text{when } \theta_0 \in \Omega_0. \qquad \square$$

Proof of Proposition 4.2. Using (2.32)–(2.33), we have:

$$\begin{aligned}
&1 - \hat{F}_N[S_0 \mid \theta] \leqslant \tilde{G}_N[S_0 \mid \theta] \leqslant \hat{G}_N[S_0 \mid \theta], \ \forall \theta, \\
&1 - \hat{G}_N[S_0 \mid \theta] \leqslant \tilde{F}_N[S_0 \mid \theta] \leqslant \hat{F}_N[S_0 \mid \theta],
\end{aligned}$$

hence

$$\begin{aligned}
&\sup_{\theta \in \Omega_0} \tilde{G}_N[S_0 \mid \theta] \leqslant \sup_{\theta \in \Omega_0} \hat{G}_N[S_0 \mid \theta], \\
&1 - \sup_{\theta \in \Omega_0} \hat{G}_N[S_0 \mid \theta] = \inf_{\theta \in \Omega_0} \{1 - \hat{G}_N[S_0 \mid \theta]\} \leqslant \inf_{\theta \in \Omega_0} \tilde{F}_N[S_0 \mid \theta], \\
&\sup_{\theta \in \Omega_0} \tilde{p}_N[S_0 \mid \theta] \leqslant \sup_{\theta \in \Omega_0} \hat{p}_N[S_0 \mid \theta].
\end{aligned}$$

Furthermore, when $\theta_0 \in \Omega_0$,

$$\begin{aligned}
&\sup_{\theta \in \Omega_0} \tilde{G}_N[S_0 \mid \theta] \leqslant \alpha_1 \Rightarrow \tilde{G}_N[S_0 \mid \theta_0] \leqslant \alpha_1, \\
&\inf_{\theta \in \Omega_0} \tilde{F}_N[S_0 \mid \theta] \geqslant 1 - \alpha_1 \Rightarrow \tilde{F}_N[S_0 \mid \theta_0] \geqslant 1 - \alpha_1, \\
&\sup_{\theta \in \Omega_0} \tilde{p}_N[S_0 \mid \theta] \leqslant \alpha_1 \Rightarrow \tilde{p}_N[S_0 \mid \theta_0] \leqslant \alpha_1,
\end{aligned}$$

hence, using Proposition 2.4, for $0 \leqslant \alpha_1 \leqslant 1$ and for $0 \leqslant \alpha \leqslant 1$,

$$\begin{aligned}
\mathsf{P}[\sup\{\hat{G}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \leqslant \alpha_1] &\leqslant \mathsf{P}[\sup\{\tilde{G}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \leqslant \alpha_1] \\
&\leqslant \mathsf{P}[\tilde{G}_N[S_0 \mid \theta_0] \leqslant \alpha_1] = \frac{I[\alpha_1 N] + 1}{N + 1},
\end{aligned}$$

$$\begin{aligned}
\mathsf{P}[\sup\{\hat{G}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \leqslant \alpha_1] &\leqslant \mathsf{P}[\inf\{\tilde{F}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \geqslant 1 - \alpha_1] \\
&\leqslant \mathsf{P}[\tilde{F}_N[S_0 \mid \theta_0] \geqslant 1 - \alpha_1] = \frac{I[\alpha_1 N] + 1}{N + 1},
\end{aligned}$$

$$\begin{aligned}
\mathsf{P}[\sup\{\hat{p}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \leqslant \alpha] &\leqslant \mathsf{P}[\sup\{\tilde{p}_N[S_0 \mid \theta] : \theta \in \Omega_0\} \leqslant \alpha] \\
&\leqslant \mathsf{P}[\tilde{p}_N[S_0 \mid \theta_0] \leqslant \alpha] = \frac{I[\alpha(N + 1)]}{N + 1}. \qquad \square
\end{aligned}$$

Proof of Proposition 5.1. Using arguments similar to the ones in the proof of Proposition 4.2 [see (A.1)–A.2)], it is easy to see that

$$\begin{aligned}\mathsf{P}[\sup\{\hat{G}_{TN}[S_{T0} \mid \theta] : \theta \in C_T\} \leqslant \alpha_1] &\leqslant \mathsf{P}[\inf\{\hat{F}_{TN}[S_{T0} \mid \theta] : \theta \in C_T\} \geqslant 1-\alpha_1] \\ &= \mathsf{P}[S_{T0} \geqslant \sup\{\hat{F}_{TN}^{-1}[1-\alpha_1 \mid \theta] : \theta \in C_T\}].\end{aligned}$$

Further,

$$\begin{aligned}&\mathsf{P}[\inf\{\hat{F}_{TN}[S_{T0} \mid \theta] : \theta \in C_T\} \geqslant 1-\alpha_1] \\ &\quad = \mathsf{P}[\inf\{\hat{F}_{TN}[S_{T0} \mid \theta] : \theta \in C_T\} \geqslant 1-\alpha_1 \text{ and } \theta_0 \in C_T] \\ &\qquad + \mathsf{P}[\inf\{\hat{F}_{TN}[S_{T0} \mid \theta] : \theta \in C_T\} \geqslant 1-\alpha_1 \text{ and } \theta_0 \notin C_T] \\ &\quad \leqslant \mathsf{P}[\hat{F}_{TN}[S_{T0} \mid \theta_0] \geqslant 1-\alpha_1] + \mathsf{P}[\theta_0 \notin C_T] = \frac{I[\alpha_1 N]+1}{N+1} + \mathsf{P}[\theta_0 \notin C_T],\end{aligned}$$

where the last identity follows from Proposition 2.2, hence, since $\lim_{T\to\infty} \mathsf{P}[\theta_0 \notin C_T] = 0$,

$$\begin{aligned}\lim_{T\to\infty} \mathsf{P}[\inf\{\hat{F}_{TN}[S_{T0} \mid \theta] : \theta \in C_T\} \geqslant 1-\alpha_1] &\leqslant \frac{I[\alpha_1 N]+1}{N+1} + \lim_{T\to\infty} \mathsf{P}[\theta_0 \notin C_T] \\ &= \frac{I[\alpha_1 N]+1}{N+1},\end{aligned}$$

from which (5.9) and (5.10) follow. □

Proof of Proposition 5.2. The result follows from arguments similar to the ones used in the proofs of Propositions 4.1 and 5.2 (with Ω_0 replaced by C_T). □

In order to prove Proposition 6.1, it will be convenient to first demonstrate the following two lemmas

Lemma A.1 (*Continuity of p-value function*). *Under the assumptions and notations* (5.1), (5.2), (5.4), (5.5), (5.14)–(5.16) *and* (6.1), *set*

$$Q_{TN}(\theta, x, u_0, \alpha_1) = \mathsf{P}[\tilde{G}_{TN}[x \mid \theta] \leqslant \alpha_1 \mid U_0 = u_0], \tag{A.3}$$

$$\bar{Q}_{TN}(\theta, x, \alpha_1) = \mathsf{P}[\hat{G}_{TN}[x \mid \theta] \leqslant \alpha_1], \quad 0 \leqslant \alpha_1 \leqslant 1, \tag{A.4}$$

and suppose U_0 *is independent of* $S_T(N, \theta)$. *For any* $\theta, \theta_0 \in \Omega$, $x \in \mathbb{R}$ *and* $u_0, \alpha_1 \in [0, 1]$, *the inequality*

$$|F_T[y \mid \theta] - F_T[y \mid \theta_0]| \leqslant \varepsilon, \quad \forall y \in (x-\delta,\ x+\delta),$$

where $\delta > 0$, *entails the inequalities*:

$$|Q_{TN}(\theta, x, u_0, \alpha_1) - Q_{TN}(\theta_0, x, u_0, \alpha_1)| \leqslant 3C(N, \alpha_1)\varepsilon, \tag{A.5}$$

$$|\bar{Q}_{TN}(\theta, x, \alpha_1) - \bar{Q}_{TN}(\theta_0, x, u_0)| \leqslant 3C(N, \alpha_1)\varepsilon, \tag{A.6}$$

where $C(N, \alpha_1) = N\sum_{k=0}^{I[\alpha_1 N]} \binom{N}{k}$.

Proof. It is easy to see [as in (3.3)] that

$$
\begin{aligned}
Q_{TN}(\theta, x, u_0, \alpha_1) &= \mathsf{P}[\tilde{G}_{TN}(x \mid \theta) \leqslant \alpha_1 \mid U_0 = u_0] \\
&= \sum_{k=0}^{I[\alpha_1 N]} \binom{N}{k} \bar{G}_T(x, u_0 \mid \theta)^k [1 - \bar{G}_T(x, u_0 \mid \theta)]^{N-k},
\end{aligned}
$$

where

$$
\begin{aligned}
\bar{G}_T(x, u_0 \mid \theta) &= \mathsf{P}(\mathbf{1}[(S_{Ti}(\theta), U_i) \geqslant (x, u_0)] = 1) \\
&= \mathsf{P}[S_{Ti}(\theta) > x] + \mathsf{P}[S_{Ti}(\theta) = x]\mathsf{P}[U_i \geqslant u_0] \\
&= 1 - F_T[x \mid \theta] + g_T(x \mid \theta)(1 - u_0), \quad 1 \leqslant i \leqslant T.
\end{aligned}
$$

Note also that $g_T(x \mid \theta) = F_T[x \mid \theta] - \lim_{\delta_0 \to 0^+} F_T[y - \delta_0 \mid \theta]$. Then the inequality

$$
|F_T[y \mid \theta] - F_T[y \mid \theta_0]| \leqslant \varepsilon, \quad \forall y \in (x - \delta, x + \delta),
$$

entails the following inequalities:

$$
|1 - F_T[x \mid \theta] - 1 + F_T[x \mid \theta_0]| \leqslant \varepsilon,
$$

$$
\begin{aligned}
|g_T(x \mid \theta) - g_T(x \mid \theta_0)| &= |\{1 - F_T[x \mid \theta]\} - \{1 - F_T[x \mid \theta_0]\} \\
&\quad + \lim_{\delta_0 \to 0^+} \{F_T[y - \delta_0 \mid \theta] - F_T[y - \delta_0 \mid \theta_0]\}| \\
&\leqslant |F_T[x \mid \theta] - F_T[x \mid \theta_0]| + \lim_{\delta_0 \to 0^+} |F_T[y - \delta_0 \mid \theta] \\
&\quad - F_T[y - \delta_0 \mid \theta_0]| \leqslant 2\varepsilon,
\end{aligned}
$$

hence, for all $u_0 \in [0, 1]$,

$$
\begin{aligned}
|\bar{G}_T(x, u_0 \mid \theta) - \bar{G}_T(x, u_0 \mid \theta_0)| &\leqslant |F_T[x \mid \theta] - F_T[x \mid \theta_0]| \\
&\quad + |1 - u_0||g_T(x \mid \theta) - g_T(x \mid \theta_0)| \leqslant 3\varepsilon, \\
&\quad \forall u_0 \in [0, 1],
\end{aligned}
$$

$$
\begin{aligned}
&|Q_{TN}(\theta, x, u_0, \alpha_1) - Q_{TN}(\theta_0, x, u_0, \alpha_1)| \\
&\leqslant \left| \sum_{k=0}^{[N\alpha_1]} \left\{ \binom{N}{k} \bar{G}_T(x, u_0|\theta)^k [1 - \bar{G}_T(x, u_0|\theta)]^{N-k} \right.\right. \\
&\quad \left.\left. - \bar{G}_T(x, u_0|\theta_0)^k [1 - \bar{G}_T(x, u_0|\theta_0)]^{N-k} \right\} \right| \\
&\leqslant \sum_{k=0}^{[N\alpha_1]} \binom{N}{k} |\bar{G}_T(x, u_0 \mid \theta)^k - \bar{G}_T(x, u_0 \mid \theta_0)^k| \\
&\quad + |[1 - \bar{G}_T(x, u_0 \mid \theta)]^{N-k} - [1 - \bar{G}_T(x, u_0 \mid \theta_0)]^{N-k}|
\end{aligned}
$$

$$\leqslant \sum_{k=0}^{[N\alpha_1]} \left\{ \binom{N}{k} k|\bar{G}_T(x, u_0 \mid \theta) - \bar{G}_T(x, u_0 \mid \theta_0)| \right.$$

$$\left. + (N-k)|\bar{G}_T(x, u_0 \mid \theta) - \bar{G}_T(x, u_0 \mid \theta_0)| \right\}$$

$$= C(N, \alpha_1)|\bar{G}_T(x, u_0 \mid \theta) - \bar{G}_T(x, u_0 \mid \theta_0)| \leqslant 3C(N, \alpha_1)\varepsilon,$$

where $C(N, \alpha_1) = N\sum_{k=0}^{I[\alpha_1 N]} \binom{N}{k}$, from which (A.5) follows. The inequality (A.6) follows in a similar way on noting that $\bar{Q}_{TN}(\theta, x, \alpha_1) = \sum_{k=0}^{I[\alpha_1 N]} \binom{N}{k} G_T(x \mid \theta)^k [1 - G_T(x \mid \theta)]^{N-k}$, where $G_T(x \mid \theta) = \mathsf{P}_\theta[S_{Ti}(\theta) \geqslant x]$, $1 \leqslant i \leqslant N$. $\square$

Lemma A.2 (*Convergence of Bootstrap p-values*). *Under the assumptions and notations of Lemma* A.1, *suppose that* (6.2) *and* (6.3) *also hold. If* $\hat{\theta}_T \underset{T\to\infty}{\longrightarrow} \theta_0$ *in probability and condition* (6.4) *or* (6.5)–(6.7) *holds*, *then*

$$\sup_{0 \leqslant u_0 \leqslant 1} |Q_{TN}(\hat{\theta}_T, S_{T0}, u_0, \alpha_1) - Q_{TN}(\theta_0, S_{T0}, u_0, \alpha_1)| \xrightarrow[T\to\infty]{p} 0, \tag{A.7}$$

$$\bar{Q}_{TN}(\hat{\theta}_T, S_{T0}, \alpha_1) - \bar{Q}_{TN}(\theta_0, S_{T0}, \alpha_1) \xrightarrow[T\to\infty]{p} 0. \tag{A.8}$$

We can now prove the following proposition.

Proof. Let $\alpha_1 \in [0, 1]$, $\varepsilon > 0$ and $\varepsilon_0 > 0$ and suppose first that (6.4) holds. Then, using Lemma A.1, we can find $\delta > 0$ and T_1 such that

$$\begin{aligned} &x \in D_{T0}(\varepsilon_0),\ \|\theta - \theta_0\| \leqslant \delta \text{ and } T > T_1 \\ &\quad \Rightarrow |F_T[x \mid \theta] - F_T[x \mid \theta_0]| \leqslant \varepsilon_1 \equiv \varepsilon/[3C(N, \alpha_1)] \\ &\quad \Rightarrow |Q_{TN}(\theta, x, u_0, \alpha_1) - Q_{TN}(\theta_0, x, u_0, \alpha_1)| \leqslant \varepsilon, \quad \forall u_0 \in [0, 1]. \end{aligned}$$

Thus

$$S_{T0} \in D_{T0}(\varepsilon_0) \text{ and } \|\hat{\theta}_T - \theta_0\| \leqslant \delta \Rightarrow \Delta_{TN}(\hat{\theta}_T, \theta_0, S_{T0}, \alpha_1) \leqslant \varepsilon,$$

where $\Delta_{TN}(\hat{\theta}_T, \theta_0, S_{T0}, \alpha_1) \equiv \sup_{0 \leqslant u_0 \leqslant 1} |Q_{TN}(\hat{\theta}_T, S_{T0}, u_0, \alpha_1) - Q_{TN}(\theta_0, x, u_0, \alpha_1)|$, hence

$$\begin{aligned} \mathsf{P}[\Delta_{TN}(\hat{\theta}_T, \theta_0, S_{T0}, \alpha_1) \leqslant \varepsilon] &\geqslant \mathsf{P}[S_{T0} \in D_{T0}(\varepsilon_0) \text{ and } \|\hat{\theta}_T - \theta_0\| \leqslant \delta] \\ &\geqslant 1 - \mathsf{P}[S_{T0} \notin D_{T0}(\varepsilon_0)] - \mathsf{P}[\|\hat{\theta}_T - \theta_0\| > \delta] \\ &= \mathsf{P}[S_{T0} \in D_{T0}(\varepsilon_0)] - \mathsf{P}[\|\hat{\theta}_T - \theta_0\| > \delta]. \end{aligned}$$

Since $\hat{\theta}_T \xrightarrow{p} \theta_0$, it follows that

$$\liminf_{T\to\infty} \mathsf{P}[\Delta_{TN}(\hat{\theta}_T, \theta_0, S_{T0}, \alpha_1) \leqslant \varepsilon] \geqslant \liminf_{T\to\infty} \mathsf{P}[S_{T0} \in D_{T0}(\varepsilon_0)] \geqslant 1 - \varepsilon_0$$

for any $\varepsilon_0 > 0$, hence $\lim_{T\to\infty} \mathsf{P}[\Delta_{TN}(\hat{\theta}_T, \theta_0, S_{T0}, \alpha_1) \leqslant \varepsilon] = 1$. Since the latter identity holds for any $\varepsilon > 0$, (A.7) is established. (A.8) follows in a similar way upon using (A.6).

Suppose now (6.5)–(6.7) hold instead of (6.4). Then, $(S_{T0}, \hat{\theta}_T) \xrightarrow[T\to\infty]{p} (S_0, \theta_0)$ and

$$(S_{\bar{T}_k}, \hat{\theta}_{\bar{T}_k}) \xrightarrow[T\to\infty]{p} (S_0, \theta_0), \tag{A.9}$$

for any subsequence $\{(S_{\bar{T}_k}, \hat{\theta}_{\bar{T}_k}) : k = 1, 2, ...\}$ of $\{(S_T, \hat{\theta}_T) : T \geqslant I_0\}$. Since S_{T0} and $\hat{\theta}_T$, $T \geqslant I_0$, are random variables (or vectors) defined on $\mathscr{X}$, we can write $S_{T0} = S_{T0}(\omega)$, $\hat{\theta}_T = \hat{\theta}_T(\omega)$ and $S_0 = S_0(\omega)$, $\omega \in \mathscr{X}$. By (6.6), the event

$$A_0 = \{\omega : S_0(\omega) \in D_0 \text{ and } S_{T0}(\omega) \in D_0, \text{ for } T \geqslant I_0\}$$

has probability one. Furthermore, by (A.9), the subsequence $(S_{\bar{T}_k}, \hat{\theta}'_{\bar{T}_k})'$ contains a further subsequence $(S_{T_k 0}, \hat{\theta}'_{T_k})'$, $k \geqslant 1$ such that $(S_{T_k 0}, \hat{\theta}'_{T_k})' \xrightarrow[T\to\infty]{} (S_0, \theta'_0)'$ *a.s.* (where $T_1 < T_2 < \cdots$); see Bierens (1994, pp. 22–23). Consequently, the set

$$C_0 = \{\omega \in \mathscr{X} : S_0(\omega) \in D_0, \ \lim_{k\to\infty} S_{T_k 0}(\omega) = S_0(\omega) \text{ and } \lim_{k\to\infty} \hat{\theta}_{T_k}(\omega) = \theta_0\}$$

has probability one. Now, let $\varepsilon > 0$. By (6.7), for any $x \in D_0$, we can find $\delta(x, \varepsilon) > 0$, $T(x, \varepsilon) > 0$ and an open neighborhood $B(x, \varepsilon)$ of x such that

$$\|\theta - \theta_0\| \leqslant \delta(x, \varepsilon) \text{ and } T > T(x, \varepsilon) \Rightarrow |F_T[y \mid \theta] - F_T[y \mid \theta_0]| \leqslant \varepsilon, \quad \forall y \in B(x, \varepsilon) \cap D_0.$$

Furthermore, for $\omega \in C_0$, we can find k_0 such that

$$k \geqslant k_0 \Rightarrow S_{T_k 0}(\omega) \in B(S_0(\omega), \varepsilon) \cap D_0 \text{ and } \|\hat{\theta}_{T_k} - \theta_0\| \leqslant \delta(S_0(\omega), \varepsilon),$$

so that $T_k > \max\{T(S_0(\omega), \varepsilon), T_{k_0}\}$ entails $|F_{T_k}[S_{T_k 0}(\omega) \mid \theta_{T_k}(\omega)] - F_{T_k}[S_{T_k 0}(\omega) \mid \theta_0]| \leqslant \varepsilon$. Thus $\lim_{k\to\infty} \{F_{T_k}[S_{T_k 0}(\omega) \mid \theta_{T_k}(\omega)] - F_{T_k}[S_{T_k 0}(\omega) \mid \theta_0]\} = 0$ for $\omega \in C_0$, hence, using Lemma A.1, $\lim_{k\to\infty} \Delta_{T_k N}(\hat{\theta}_{T_k}(\omega), \theta_0, S_{T_k 0}(\omega), \alpha_1) = 0$ and $\Delta_{T_k N}(\hat{\theta}_{T_k}, \theta_0, S_{T_k 0}, \alpha_1) \xrightarrow[k\to 0]{} 0$, a.s. This shows that any subsequence of the sequence $\Delta_{TN}(\hat{\theta}_T, \theta_0, S_{T0}, \alpha_1), T \geqslant I_0$, contains a further subsequence which converge a.s. to zero. It follows that $\Delta_{TN}(\hat{\theta}_T, \theta_0, S_{T0}, \alpha_1) \xrightarrow[T\to\infty]{p} 0$ and (A.7) is established. The proof of (A.8)) under the condition (6.5)–(6.7) is similar. □

Proof of Proposition 6.1. Using the fact that $\hat{\theta}_T$, S_{T0} and U_0 are independent of $S_T(N, \theta)$, we can write

$$\begin{aligned}
&\mathsf{P}[\tilde{G}_{TN}[S_{T0} \mid \hat{\theta}_T] \leqslant \alpha_1] - \mathsf{P}[\tilde{G}_{TN}[S_{T0} \mid \theta_0] \leqslant \alpha_1] \\
&\quad = \mathsf{E}\{\mathsf{P}[\tilde{G}_{TN}[S_{T0} \mid \hat{\theta}_T] \leqslant \alpha_1 \mid (\hat{\theta}_T, S_{T0}, U_0)] \\
&\qquad - \mathsf{P}[\tilde{G}_{TN}[S_{T0} \mid \theta_0] \leqslant \alpha_1 \mid (\hat{\theta}_T, S_{T0}, U_0)]\} \\
&\quad = \mathsf{E}[Q_{TN}(\hat{\theta}_T, S_{T0}, U_0, \alpha_1) - Q_{TN}(\theta_0, S_{T0}, U_0, \alpha_1)].
\end{aligned}$$

From Lemma A.1 and using the Lebesgue dominated convergence theorem, we then get

$$\begin{aligned}
&|\mathsf{P}[\tilde{G}_{TN}[S_{T0}|\hat{\theta}_T]\leqslant\alpha_1] - \mathsf{P}[\tilde{G}_{TN}[S_{T0}|\theta_0]\leqslant\alpha_1]| \\
&\quad = |\mathsf{E}[Q_{TN}(\hat{\theta}_T, S_{T0}, U_0, \alpha_1) - Q_{TN}(\theta_0, S_{T0}, U_0, \alpha_1)]| \\
&\quad \leqslant \mathsf{E}\{|Q_{TN}(\hat{\theta}_T, S_{T0}, U_0, \alpha_1) - Q_{TN}(\theta_0, S_{T0}, U_0, \alpha_1)|\} \\
&\quad \leqslant \mathsf{E}\left[\sup_{0\leqslant u_0\leqslant 1} |Q_{TN}(\hat{\theta}_T, S_{T0}, u_0, \alpha_1) - Q_{TN}(\theta_0, S_{T0}, u_0, \alpha_1)|\right] \underset{T\to\infty}{\longrightarrow} 0.
\end{aligned}$$

We can show in a similar way that

$$|\mathsf{P}[\hat{G}[S_{T0} \mid \hat{\theta}_T]\leqslant\alpha_1] - \mathsf{P}[G[S_{T0} \mid \theta_0]\leqslant\alpha_1]| \underset{T\to\infty}{\longrightarrow} 0,$$

from which we get (6.8). Eq. (6.9) then follows from the definitions of $\tilde{p}_{TN}(x \mid \theta)$ and $\hat{p}_{TN}(x \mid \theta)$. $\square$

References

Andrews, D.W.K., 2000. Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. Econometrica 68, 399–405.

Athreya, K.B., 1987. Bootstrap of the mean in the infinite variance case. The Annals of Statistics 15, 724–731.

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

Basawa, I.V., Mallik, A.K., McCormick, W.P., Reeves, J.H., Taylor, R.L., 1991. Bootstrapping unstable first-order autoregressive processes. The Annals of Statistics 19, 1098–1101.

Benkwitz, A., Lütkepohl, H., Neumann, M.H., 2000. Problems related to confidence intervals for impulse responses of autoregressive processes. Econometric Reviews 19, 69–103.

Beran, R., Ducharme, G. R., 1991. Asymptotic Theory for Bootstrap Methods in Statistics. Centre de Recherches Mathématiques, Université de Montréal, Montréal, Canada.

Besag, J., Diggle, P.J., 1977. Simple Monte Carlo tests for spatial pattern. Applied Statistics 26, 327–333.

Bierens, H.J., 1994. Topics in Advanced Econometrics. Cambridge University Press, New York.

Birnbaum, Z.W., 1974. Computers and unconventional test-statistics. In: Proschan, F., Serfling, R.J. (Eds.), Reliability and Biometry. SIAM, Philadelphia, PA, pp. 441–458.

Brown, L.D., Purves, R., 1973. Measurable selections of extrema. The Annals of Statistics 1, 902–912.

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

Chernick, M.R., 1999. Bootstrap Methods: A Practitioner's Guide. Wiley, New York.

Davison, A., Hinkley, D., 1997. Bootstrap Methods and Their Application. Cambridge University Press, Cambridge, UK.

Debreu, G., 1967. Integration of correspondences. In: LeCam, L., Neyman, J. (Eds.), Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. II-1. The University of California Press, Berkeley and Los Angeles, California, pp. 351–372.

Dufour, J.-M., 1989. Nonlinear hypotheses, inequality restrictions, and non-nested hypotheses: Exact simultaneous tests in linear regressions. Econometrica 57, 335–355.

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

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

Dufour, J.-M., Jouini, T., 2005. Finite-sample simulation-based tests in VAR models with applications to causality testing. Journal of Econometrics, forthcoming.

Dufour, J.-M., Khalaf, L., 2003a. Finite sample tests in seemingly unrelated regressions. In: Giles, D.E.A. (Ed.), Computer-Aided Econometrics. Marcel Dekker, New York, pp. 11–35 (Chapter 2).

Dufour, J.-M., Khalaf, L., 2003b. Simulation-based finite-sample inference in simultaneous equations. Technical report, Centre interuniversitaire de recherche en analyse des organisations (CIRANO) and Centre interuniversitaire de recherche en économie quantitative (CIREQ), Université de Montréal.

Dufour, J.-M., Kiviet, J.F., 1996. Exact tests for structural change in first-order dynamic models. Journal of Econometrics 70, 39–68.

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

Dufour, J.-M., Valéry, P., 2005. Exact and asymptotic tests for possibly non-regular hypotheses on stochastic volatility models. Technical report. Centre interuniversitaire de recherche en analyse des organisations (CIRANO) and Centre interuniversitaire de recherche en économie quantitative (CIREQ), Université de Montréal.

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

Dufour, J.-M., Farhat, A., Gardiol, L., Khalaf, L., 1998. Simulation-based finite sample normality tests in linear regressions. The Econometrics Journal 1, 154–173.

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

Edgington, E.S., 1980. Randomization Tests. Marcel Dekker, New York.

Edgington, E.S., 1995. Randomization Tests, third ed. Marcel Dekker, New York.

Edwards, D., 1985. Exact simulation-based inference: A survey, with additions. Journal of Statistical Computation and Simulation 22, 307–326.

Edwards, D., Berry, J.J., 1987. The efficiency of simulation-based multiple comparisons. Biometrics 43, 913–928.

Efron, B., 1982. The Jacknife, the Bootstrap and Other Resampling Plans CBS-NSF Regional Conference Series in Applied Mathematics, Monograph No. 38. Society for Industrial and Applied Mathematics, Philadelphia, PA.

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap of Monographs on Statistics and Applied Probability, vol. 57. Chapman & Hall, New York.

Foutz, R.V., 1980. A method for constructing exact tests from test statistics that have unknown null distributions. Journal of Statistical Computation and Simulation 10, 187–193.

Gallant, A.R., Tauchen, G., 1996. Which moments to match? Econometric Theory 12, 657–681.

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

Good, P., 1994. Permutation Tests. A Practical Guide to Resampling Methods for Testing Hypotheses, second ed. Springer, New York.

Gouriéroux, C., Monfort, A., 1996. Simulation-Based Econometric Methods. Oxford University Press, Oxford, UK.

Green, B.F., 1977. A practical interactive program for randomization tests of location. The American Statistician 31, 37–39.

Hájek, J., 1969. A Course in Nonparametric Statistics. Holden-Day, San Francisco.

Hajivassiliou, V.A., 1993. Simulation estimation methods for limited dependent variables. In: Maddala, et al. (Eds.), Handbook of Statistics 11: Econometrics. North-Holland, Amsterdam, pp. 519–543.

Hall, P., 1992. The Bootstrap and Edgeworth Expansion. Springer, New York.

Hall, P., Titterington, D.M., 1989. The effect of simulation order on level accuracy and power of Monte Carlo tests. Journal of the Royal Statistical Society, Series B 51, 459–467.

Hope, A.C.A., 1968. A simplified Monte Carlo test procedure. Journal of the Royal Statistical Society, Series B 30, 582–598.

Horowitz, J.L., 1997. Bootstrap methods in econometrics: Theory and numerical performance. In: Kreps, D., Wallis, K.W. (Eds.), Advances in Economics and Econometrics, vol. 3. Cambridge University Press, Cambridge, UK, pp. 188–222.

Inoue, A., Kilian, L., 2002. Bootstrapping autoregressive process with possible unit roots. Econometrica 70 (1), 377–391.

Inoue, A., Kilian, L., 2003. The continuity of the limit distribution in the parameter of interest is not essential for the validity of the bootstrap. Econometric Theory 19 (6), 944–961.

Jeong, J., Maddala, G.S., 1993. A perspective on application of bootstrap methods in econometrics. In: Maddala, et al. (Eds.), Handbook of Statistics 11: Econometrics. North-Holland, Amsterdam, pp. 573–610.

Jöckel, K.-H., 1986. Finite sample properties and asymptotic efficiency of Monte Carlo tests. The Annals of Statistics 14, 336–347.

Jouneau-Sion, F., Torrés, O., 2005. MMC techniques for limited dependent variables models: Implementation by the branch and bound algorithm. Journal of Econometrics, this issue, doi:10.1016/j.jeconom.2005.06.006.

Keane, M.P., 1993. Simulation estimation for panel data models with limited dependent variables. In: Maddala, et al. (Eds.), Handbook of Statistics 11: Econometrics. North-Holland, Amsterdam, pp. 545–571.

Keller-McNulty, S., Higgins, J.J., 1987. Effect of tail weight and outliers and type-I error of robust permutation tests for location. Communications in Statistics, Simulation and Computation 16 (1), 17–35.

Kiviet, J.F., Dufour, J.-M., 1997. Exact tests in single equation autoregressive distributed lag models. Journal of Econometrics 80, 325–353.

Lehmann, E.L., 1986. Testing Statistical Hypotheses, second ed. Wiley, New York.

Lock, R.H., 1991. A sequential approximation to a permutation test. Communications in Statistics, Simulation and Computation 20 (1), 341–363.

Manly, B.F.J., 1997. Randomization, Bootstrap and Monte Carlo Methods in Biology; Permutation Test, second ed. Chapman & Hall, London.

Mariano, R.S., Brown, B.W., 1993. Stochastic simulation for inference in nonlinear errors-in-variables models. In: Maddala, et al. (Eds.), Handbook of Statistics 11: Econometrics. North-Holland, Amsterdam, pp. 611–627.

Marriott, F.H.C., 1979. Barnard's Monte Carlo tests: How many simulations? Applied Statistics 28, 75–77.

McFadden, D., 1989. A method of simulated moments for estimation of discrete response models without numerical integration. Econometrica 57, 995–1026.

Noreen, E.W., 1989. Computer Intensive Methods for Testing Hypotheses: An Introduction. Wiley, New York.

Reiss, H.D., 1989. Approximate Distributions of Order Statistics with Applications to Nonparametric Statistics Springer Series in Statistics. Springer, New York.

Ripley, B.D., 1981. Spatial Statistics. Wiley, New York.

Shao, S., Tu, D., 1995. The Jackknife and Bootstrap. Springer, New York.

Shorack, G.R., Wellner, J.A., 1986. Empirical Processes with Applications to Statistics. Wiley, New York.

Sriram, T.N., 1994. Invalidity of bootstrap for critical branching processes with immigration. The Annals of Statistics 22, 1013–1023.

Stinchcombe, M.B., White, H., 1992. Some measurability results for extrema of random functions over random sets. Review of Economic Studies 59, 495–512.

Vadiveloo, J., 1983. On the theory of modified randomization tests for nonparametric hypotheses. Communications in Statistics, Theory and Methods 12 (14), 1581–1596.

Vinod, H.D., 1993. Bootstrap methods: Applications in econometrics. In: Maddala, et al. (Eds.), Handbook of Statistics 11: Econometrics. North-Holland, Amsterdam, pp. 629–661.