Inference for Stochastic Dominance Using Large Deviations

Asymptotics

Thomas Parker∗
Department of Economics, University of Waterloo
Email: tmparker@uwaterloo.ca
August 14, 2015

Abstract
A simple method for uniform inference on stochastic dominance relations of order greater than 1 is demonstrated using approximations to the supremum of a Gaussian process. Specifically, supremum-norm test statistics based on integrated empirical processes converge to suprema of limiting integrated Gaussian processes; a large deviations approximation is provided for the tail of the supremum distribution of these limiting processes, allowing uniform asymptotic inference over any bounded intervals of interest. This extends analytic results discussed in, for example, McFadden (1989) to all higher orders of stochastic dominance testing and may be useful as a complement to resampling strategies usually used for inference.

Keywords: Stochastic dominance, integrated Gaussian process, large deviations approximation

1 Introduction

Stochastic dominance relations have a special status in inference for poverty and inequality measurement, since they imply similar orderings for many indices that depend on the distribution functions from two populations. Recent advances have been made on inference for stochastic dominance relationships between two or more populations. Work by Anderson (1996) or Davidson and Duclos (2000), for example, proposed inferential methods that were computed using a finite number of points. Although practical, these methods are inconsistent because they depend on evaluation at an arbitrarily chosen set of comparison points, and therefore interest has also been focused on consistent inference, where the supremum norm of a suitable stochastic process is usually used as the test statistic. Notable research on consistent inferential methods includes McFadden (1989), which described first-order tests and proposed approximations for second order tests, and Barrett and Donald (2003), Linton et al. (2005) and Linton et al. (2010), which proposed resampling methods for sup-norm inference.

∗The author would like to thank Pierre Chaussé for helpful discussions. An earlier version of this work benefited from input from Juan Carlos Escanciano, Andreas Hagemann, Roger Koenker and Gene Savin.
The exact distribution of sup-norm statistics is in general quite difficult to calculate, if not completely intractable. For this reason Barrett and Donald (2003) and Linton et al. (2005) provided multiplier bootstrap and subsampling tests respectively, for any order of dominance relation, and Linton et al. (2005) allow the tests to consider several samples that may be dependent upon one another. Horváth et al. (2006) and Linton et al. (2010) propose improvements to tests that improve the size and power of tests under general conditions. These methods also apply when the distribution of tests is affected by parameter estimation (see Donald et al. (2012) for a review of the multiplier bootstrap method as applied to conditional inference).

Resampling for the empirical processes used in this context only relies on the existence of some well-behaved Gaussian limit process. They rely on no more than the existence of such a process which is a strength. However, it is occasionally useful to have an asymptotic approximation to make quick calculations, provide context or to compute inferential results in large samples. Indeed, the latter half of Section 4 of McFadden (1989) is concerned with finding bounds and approximations to the tail probabilities of the sup-norm of the process used to test second degree stochastic dominance, and Klecan et al. (1991), in their discussion of the lack of tractable asymptotic approximations, mention an approximation due to Adler and Samorodnitsky (1987) that resembles the result specified in Theorem 3.2 below. In this work the properties of the limiting Gaussian processes associated with tests of stochastic dominance of order 2 and higher are considered and a simple asymptotic characterization of the tail behavior of their supremum norms is derived under general conditions. This makes asymptotic inference as easy to implement as the resampling methods proposed in the papers cited above.

The inferential methods proposed here are based on a large deviations characterization of the law of the supremum norm of the aforementioned Gaussian limit processes. Because the term “large deviations” is used differently in several contexts, it is worthwhile discussing for a moment the term as it is used here and how these large deviations results developed in the probability literature. The law of the supremum of a Gaussian process is known for only a few specific processes (Adler and Taylor, 2007, Chapter 4, Footnote 2), and these results have mainly relied on convenient mathematical properties like Markov- and stationarity conditions. Therefore research into the distribution of the supremum has shifted towards large deviations asymptotics, analyzing the right tail of the distribution, which is usually rather well-behaved. One area of research has focused on logarithmic asymptotics, which are quite general results that take the form \( \lim_{u \to \infty} \log P \{ \sup_t X_t > u \} / g(u) = 1 \) as \( u \to \infty \); see, for example, Dembo and Zeitouni (1998). However, such generality comes at the cost of a lack of specificity. Another strand in the large deviations probability literature is sometimes labeled exact large deviations asymptotics, where “exact” is used to denote the fact that asymptotics are derived directly for the excursion probability \( P \{ \sup_t X_t > u \} \) rather than its logarithm; that is, results take the form \( \lim_{u \to \infty} P \{ \sup_t X_t > u \} / h(u) = 1 \). Exact asymptotics do not apply in situations as general as logarithmic asymptotics, but can provide more precise results for use in inference when the properties of the process and its parameter space are well-behaved. The monograph of Lifshits (1995, Chapters 12 and 13) illustrates the distinction between logarithmic and exact asymptotics nicely\(^1\). The [exact] asympt-
totics resulting from the double-sum technique illustrated in Piterbarg (1996), which has its roots in Pickands (1969), have found applications in such fields as signal processing and fluid dynamics. The monograph of Adler and Taylor (2007) makes fascinating connections between such large deviations results, differential geometry and tube formulas and provides asymptotics for a wide variety of (mainly stationary) stochastic processes.

The class of processes that have distributions that can be described using these techniques are those that can be expressed as integrals of other stochastic process, and as such the asymptotics here are applied to tests of second-order stochastic dominance or higher. Economic applications of second-order stochastic dominance are most often towards inequality comparisons, since a second-order dominance comparison is equivalent to comparison of the average poverty gap in two distributions of income (whereas first-order dominance, measuring the proportion of individuals below a poverty line, is a natural measure of poverty). However, other related measures include the prospect stochastic dominance concept introduced by Kahneman and Tversky (1979) and studied in Linton et al. (2005) and the concept of Markowitz stochastic dominance introduced in Levy and Levy (2002), which both employ a criterion similar to the second-order stochastic dominance criterion, but over different subdomains of the real line. For a review of the way in which second-order dominance comparisons are related to the cumulative poverty gap and generalized Lorenz curves, see Davidson and Duclos (2000, Section 2).

A few caveats should be made regarding the limitations of the approach below. First, it is asymptotic in two senses of the word — as levels $u \to \infty$ and as sample sizes $n \to \infty$ — so it is possible that the finite-sample behavior of tests based on the supremum of empirical processes will not be accurately captured. Some smoothness of the distribution function must also be assumed, which may be a drawback in applications that have unusual features such as discontinuities at points of interest. Finally, tests of first-order stochastic dominance, perhaps the most widely-used of all such tests, can be quite different from processes associated with all higher-order tests, and it is much more difficult to accurately capture the distribution of the supremum of these processes. Large-deviations results are provided below for processes used in tests that have no parameter estimation, but it is not generally possible to provide such results for the case with estimated parameters when tests are for first-order stochastic dominance. Resampling offers a tractable solution to these problems.

In the next section the basic hypotheses and statistics used for stochastic dominance inference are established and discussed. Section 3 discusses the main theoretical results for a class of integrated Gaussian processes defined there. These results are then applied to the tests in question in Section 4. Proof of the theorems in the text is given in an appendix.

2 Empirical processes for consistent tests of dominance

Suppose that $X_1$ and $X_2$ are continuous, real-valued random variables that come from distributions that may represent incomes or returns in two different areas or time periods, with marginal distribution for the continuous processes considered here, but can be manipulated to apply to countable parameter sets and the $L_p$ norm of some continuous processes.
functions $F_1$ and $F_2$.

For any cumulative distribution function $F$ with support such that $T \subseteq \text{supp}(F)$ define the following family of functionals for $j \geq 1$ and $t \in T$:

$$I_j(t, F) = \begin{cases} F(t) & j = 1 \\ \int_0^t I_{j-1}(s, F)ds & j = 2, 3, \ldots \end{cases}.$$ (1)

Using integration by parts the above functionals may also be written as

$$I_j(t, F) = \int_0^t \frac{(t-s)^{j-1}}{(j-1)!} dF(s).$$ (2)

The lower endpoint could be different, although note that for other lower endpoints (besides 0 or $-\infty$), this definition is no longer generally equal to definition 1. For any order $j \geq 1$, $X_1$ stochastically dominates $X_2$ at order $j$ on $T$ if $I_j(t, F_1) \leq I_j(t, F_2)$ for all $t \in T$. The hypothesis that $X_2$ dominates $X_1$ stochastically at the $j^{th}$ order can be tested using the formal hypotheses

$$H_0^j : I_j(t, F_2) \leq I_j(t, F_1) \quad \text{for all } t \in T \quad \text{(dominance)}$$

$$H_1^j : I_j(t, F_2) > I_j(t, F_1) \quad \text{for some } t \in T \quad \text{(nondominance)}$$

Older testing procedures such as Davidson and Duclos (2000) may suffer from inconsistency due to the arbitrary set of point at which the above hypotheses must be evaluated. Since the work of Barrett and Donald (2003), most papers in the econometric literature on testing for stochastic dominance have focused on consistent tests, that is, tests for which the probability of rejection of the null tends to 1 as sample sizes become large. These rely on some metric of an empirical process; here we consider the supremum metric, as many other authors have before.

Suppose that one is interested in testing the hypothesis that one random variable dominates another stochastically at an order $j \geq 2$, and that this judgement is to be made uniformly over an interval $t \in T = [T, \bar{T}]$, with $-\infty < T \leq \bar{T} < \infty$ using the supremum norm — that is, rejecting $H_0^j$ when values of sup$_{t \in T} (I_j(t, F_2) - I_j(t, F_1))$ are significantly greater than zero. This “restricted” sense of comparison is used in several other papers including Barrett and Donald (2003) and Davidson and Duclos (2013); the latter (in their Section 5) make a thorough argument for considering restricted comparisons. Such intervals arise naturally in inequality comparisons, for example, where second-order stochastic dominance tests are of interest. Recall that in the context of income inequality comparisons, when $X_1$ stochastically dominates $X_2$ up to a point $t$ (for example a poverty line), then the average poverty gap for individuals represented by $X_1$ is smaller than that of $X_2$. Checking second-order dominance uniformly over $T$ ensures that one’s inference with regard to such poverty comparisons is not influenced an arbitrarily chosen poverty line. When considering distributions of financial returns, positive and negative regions of the line may be of interest and can be considered separately, with 0 as an endpoint of different subintervals (also the case for applications to prospect- or Markowitz dominance hypotheses).
Horváth et al. (2006) treat general conditions necessary to work with distributions defined over all of \( \mathbb{R} \), which require weighting functions that taper to zero as \( t \to \pm \infty \), and pairs of weight functions are considered in Linton et al. (2010). Considering only bounded regions \( T \) in this work is equivalent to using a truncation-type weight function rather than a weight function that tapers off for extreme values. General asymptotic statements appear intractable in such cases, since they would typically need to be tailored to the weight function and use some knowledge of the data distribution, and resampling methods are probably to be preferred. In any case, care must be taken with tests on unrestricted intervals since without proper weighting the processes used for testing may have unbounded distributions, giving their sup-norms degenerate distributions.

The family of transformations (1) are typically used to test stochastic dominance of any order using empirical processes since (2) makes it plain that they are linear transformations of the distribution function \( F \). Specifically, let \( F_1 \) and \( F_2 \) be empirical distribution functions constructed from samples of size \( m \) and \( n \) from each population, and assume that \( \lim_{m,n \to \infty} \frac{m}{n+m} = \lambda \in (0,1) \). Define the integrated empirical process, referred to below as the \( j \)-th-order empirical process, by

\[
V_{jm}(t) = \sqrt{\frac{mn}{m+n}} \left\{ I_j(t,F_2) - I_j(t,F_1) \right\}, \quad t \in T. \tag{3}
\]

To discuss the asymptotic behavior of \( V_{jm} \), define the contact set \( S = \{ t \in \mathbb{R} : I_j(t,F_2) = I_j(t,F_1) \} \). One major concern addressed in Linton et al. (2005, Theorem 1) and Linton et al. (2010, Theorem 1) was the weak convergence result \( V_{jm} \sim V_j \), where \( V_j \) is a continuous Gaussian process, and by the continuous mapping theorem, \( \sup_{t \in T} V_{jm}(t) \sim \sup_{t \in S \cap T} V_j(t) \) where by convention the supremum of an empty set is \(-\infty\). The regularity conditions they use are weak enough to allow the processes to be dependent on one another, and their distributions may be affected by parameter estimation (even nonparametric estimates in the case of Linton et al. (2010)).

It is well known that when no parameters are estimated, \( \sqrt{n}(F_{kn} - F_k) \sim B_{F_k}, \quad k = 1,2 \), where \( B_{F_k} \) are time-changed Brownian bridges; that is, satisfying \( B_{F_k}(t) = B \circ F_k(t) \) where \( B \) is a standard Brownian bridge process on \([0,1]\). When samples are independent and under the least-favorable case of the null hypothesis, where \( F_1 \equiv F_2 \equiv F \), the sup-norm statistic \( \sup_{t \in \mathbb{R}} V_{1mn}(t) \) converges to a statistic with a distribution that is identical to that of the standard Kolmogorov-Smirnov test statistic, for which inference is readily facilitated through the formula

\[
P \left\{ \sup_{t \in \mathbb{R}} V_1(t) > u \right\} = P \left\{ \sup_{t \in \mathbb{R}} B_{F}(t) > u \right\} = e^{-2u^2}. \tag{4}
\]

As emphasized above, this limit result only holds under the very special conditions that samples are independent, distribution functions are identical, and the domain is the real line. Furthermore, for orders \( j \geq 2 \) the limiting processes have nonstandard distributions that may depend on features of \( F \), and no exact analytic solution is apparent. Therefore, inference has typically been based on simulation methods.
3 Large deviations for suprema of integrated Gaussian processes

The simple fact that the processes used for testing are themselves integrals of other processes implies that the main result below, Theorem 3.2 requires very few assumptions. In particular, it is only assumed that the first-order process is a continuous Gaussian process. The fact that the other processes in question are integrated implies that the variance is monotonically increasing and that the process is differentiable, which in turn implies (cf. Theorem 3.2) that over a finite interval \( T \), the distribution of the supremum of the process over \( T \) is asymptotically (in level) the same as the distribution of the process evaluated at the right endpoint of \( T \).

The definition below describes the specific class of processes that are treated in the subsequent theorems and their domain of definition, which is always assumed to be a closed interval.

**Definition 1** (Integrated Gaussian process). Let the parameter set \( T = [T, \overline{T}] \) be a closed interval in \( \mathbb{R} \). Suppose \( \xi : T \to \mathbb{R} \) is a Gaussian process defined by \( \xi_t = \int_T^t \xi'_s ds \) for all \( t \in T \), where the derivative process \( \xi' \) is a continuous mean-zero Gaussian process. Assume that \( \xi' \) has positive variance except perhaps at the endpoints of \( T \). Then \( \xi \) is an integrated Gaussian process on \( T \).

Note that this definition is not the most general class of what might be labeled “integrated Gaussian processes” due to the restrictions on the parameter set \( T \) and the derivative process. However, no restrictions are made with respect to the stationarity of the process or its derivative process.

Each member of this family of processes is almost surely absolutely continuous by definition. Each is also mean-square differentiable, because the fact that \( \xi' \) is a conventional derivative function of \( \xi \) implies that \( (\xi_{t+h} - \xi_t)/h \to \xi'_t \) as \( h \to 0 \) for any \( t \in T \). A good deal is known about processes that are mean-square continuous and/or m.-s. differentiable, and many of these properties are closely related to properties that their covariance functions possess. Throughout, \( \rho : T \times T \to \mathbb{R} \) is used to denote the covariance function of the process \( \xi \). Most importantly, the covariance function of a mean-square differentiable process is twice differentiable in its arguments (Loéve, 1978, Theorem C of §37.2, p. 136).

Furthermore, \( \frac{\partial^2}{\partial s \partial t} \rho(s, t) = \text{Cov}(\xi'_s, \xi'_t) \) (note that \( \rho \) is a symmetric function) and

\[
\frac{\partial^2}{\partial s \partial t} \rho(s, t) = \text{Cov}(\xi'_s, \xi'_t). \tag{5}
\]

These properties help to develop an intuition for the results below.

**Lemma 3.1.** Suppose \( \xi \) is an integrated Gaussian process on \( T \). Then

\[
\overline{T} = \arg\max_{t \in T} \text{Var}(\xi_t). \tag{6}
\]

Given this preliminary result, the next theorem shows that the upper tail of the statistic \( \sup_{t \in T} \xi_t \) has approximately the same distribution as the upper tail of the distribution of the random variable \( \xi_{\overline{T}} \); that is, the process evaluated at the right endpoint of \( T \). This is due to the fact that these processes are very smooth; informally, this smoothness is equivalent to autocorrelation that is strong enough that points away from \( \overline{T} \) contribute a negligible amount to the distribution of the supremum. This is not the
case for the Brownian bridge, for example, which exhibits locally stationary behavior near its point of maximal variance that affects the supremum law — see the discussion at the end of this section.

In all of the results below, the function $\Psi$ refers to the upper tail probability of the standard normal distribution; that is,

$$
\Psi(u) = \int_u^\infty \frac{e^{-x^2/2}}{\sqrt{2\pi}} \, dx.
$$

(7)

**Theorem 3.2.** Suppose $\xi$ is an integrated Gaussian process on $\mathcal{T}$. Then

$$
\lim_{u \to \infty} \frac{P\{\sup_{t \in \mathcal{T}} \xi_t > u\}}{\Psi\left(u/\sigma_T\right)} = 1,
$$

(8)

where $\sigma_T^2 = \text{Var}(\xi_T)$.

The proof of Theorem 3.2 is split into two parts, corresponding to whether the variance of $\xi'$ at $\mathcal{T}$ is positive or tends to zero at $\mathcal{T}$. For cases where the variance is positive, proof could be derived from Theorem 8.2 (or Theorem D.3, with a few slight modifications) of Piterbarg (1996). However, this method breaks down when $\lim_{t \uparrow \mathcal{T}} \text{Var}(\xi'_t) = 0$ because it relies on Taylor expansions of the standard deviation and correlation functions of the process. When $\text{Var}(\xi'_T) = 0$, the expansion of the correlation function would need to rely on third derivatives of the covariance function, which may not generally exist — specifically, the covariance function of a second-order process may be only twice differentiable. Therefore the proof of the above theorem relies on the Theorem of Talagrand (1988), which shows that the above result is equivalent to the behavior of the covariance function in a shrinking neighborhood of the point of maximal variance. The most important examples of an integrated Gaussian process with derivatives that tend to zero are second-order processes (i.e., once-integrated) where $\mathcal{T}$ is the upper limit of the support of the distribution of the processes, since the derivative function may tend to zero in probability there. The simplest example of this phenomenon is the once-integrated standard Brownian bridge on the unit interval, assuming the analyst is interested in the distribution of the supremum over the entire interval. This example will be illustrated in a Corollary below.

The method of proof for Theorem 3.2 exploits the characterization of the Gaussian process in question as an absolutely continuous process with a derivative function that is another Gaussian process. A perhaps more elegant method would involve entropy integrals to bound the distribution of the supremum of the process in certain neighborhoods, as in the proof of Lemma A.4 of Linton et al. (2010) (cf. the proof of their Theorem 1). However, using the entropy bounds that are sufficient to show weak convergence, but not necessary for nonstationary processes, does not result in inequalities that are strong enough to imply the above result. Further research on the class of functions used to characterize the iterated integrals may reveal tighter entropy bounds for a larger class of processes used for inequality comparisons.

Theorem 3.2 implies some results that may be of interest independent of the applications to testing discussed in the next sections. For example, consider the case of standard Brownian motion on $\mathcal{T} := [0, \mathcal{T}]$ — that is, the Gaussian process $W$ with $E[W_s] \equiv 0$ and $\text{Cov}(W_s, W_t) = s \wedge t (= \min\{s, t\})$ for all $s, t \in \mathcal{T}$. The covariance function of all higher-order processes can be calculated explicitly using
stochastic integration and calculations such as those in the proof of Lemma 1 of Barrett and Donald (2003):

\[
\text{Cov}(W^j_i, W^j_t) = \sum_{\ell=0}^{j-1} \binom{2j-\ell-2}{j-1} |s-t|^\ell (s \wedge t)^{2j-\ell-1} \frac{\ell!}{(2j-\ell-1)!}.
\]  

(9)

Theorem 3.2 implies the following Corollary for this family of integrated Brownian motions.

**Corollary 3.3.** Let \( W^j \) be the \( j \)-th-order \((j \geq 2)\) integrated Brownian motion on \([0, T]\) — that is, the Gaussian process on \([0, T]\) with mean zero and covariance function (9). Then for large \( u \),

\[
P\left\{ \sup_{t \in [0, T]} W^j_t > u \right\} \approx \Psi\left( \frac{(j-1)! \sqrt{2j-1}}{T^{j-1/2}} u \right).
\]  

(10)

Corollary 3.3 agrees with equation (1.5) of Gao and Yang (2015), which investigated integrated Brownian motions on the unit interval.

Similarly, a system of approximations is fairly straightforward to derive for the family of integrated Brownian bridges. Suppose \( B^j \) is the \( j \)-th-order integrated Brownian bridge on \([0, T]\) — that is, for \( j = 1 \), \( B^1 \) is the Gaussian process on \([0, T]\) with mean zero and covariance \( \text{Cov}(B^1_s, B^1_t) = s \wedge t - \frac{u^2}{2} \). Using, for example, the characterization \( B_t = W_t - \frac{1}{2}s W_T \), where \( W \) is the Brownian motion defined above, it can be similarly shown that the covariance of the \( j \)-th-order integrated Brownian bridge process is

\[
\text{Cov}(B^j_s, B^j_t) = \sum_{\ell=0}^{j-1} \binom{2j-\ell-2}{j-1} |s-t|^\ell (s \wedge t)^{2j-\ell-1} \frac{\ell!}{(2j-\ell-1)!} \frac{1}{T^{j-1/2}}.
\]  

(11)

Theorem 3.2 applies to this family of processes as well, and results in the following Corollary.

**Corollary 3.4.** Let \( B^j \) be the \( j \)-th-order \((j \geq 2)\) integrated Brownian bridge on \([0, T]\) — that is, the Gaussian process on \([0, T]\) with mean zero and covariance function (11). Then for large \( u \), for any interval \( S = [s, S] \subseteq [0, T] \),

\[
P\left\{ \sup_{t \in S} B^j_t > u \right\} \approx \Psi\left( \frac{j! \sqrt{T(2j-1)}}{S^{j-1/2}} \frac{T(2j-1)}{T j^2 - (2j-1) S} u \right).
\]  

(12)

When \( S = T \), the expression in the above Corollary simplifies to

\[
P\left\{ \sup_{t \in S} B^j_t > u \right\} \approx \Psi\left( \frac{j! \sqrt{2j-1}}{(j-1) T^{j-1/2}} u \right).
\]  

(13)

These results match those of Table 1 of Schmid and Trede (1998) fairly closely; they simulated standard Brownian bridges to produce quantiles of the distribution of the integrated Brownian bridge, and their simulation results match the theoretical tail quantile approximation \( Q_{B^j}(\tau) \approx \Phi^{-1}(\tau)/(2\sqrt{3}) \) (where \( \Phi^{-1} \) is the quantile function of the standard normal distribution) to ever-increasing accuracy as the quantile level increases. This is depicted in Table 1.

Note that the supremum distributions of the above integrated Brownian processes are rather different from those of the first order processes. On the same interval \([0, T]\), classical results describing
the supremum of Brownian motions and bridges assert that the order-1 processes have the following distributions, written in a manner to make it as comparable to the above results as possible:

$$P\left\{ \sup_{t \in [0,T]} W_1^1 > u \right\} = 2\Psi(u/\sqrt{T})$$

(14)

$$P\left\{ \sup_{t \in [0,T]} B_1^1 > u \right\} \approx \sqrt{2\pi} \frac{2u}{\sqrt{T}} \Psi(2u/\sqrt{T}).$$

(15)

These processes are fundamentally rougher, which makes the distribution of their suprema noticeably different from the integrated processes discussed above. In both cases (i.e., when comparing (14) to (10) or when comparing (15) to (13)) the coefficient scaling the normal tail approximation is different, and for the Brownian bridge case the variance inside the normal tail function is significantly different.

4 Uniform inference for stochastic dominance relationships

Theorem 3.2 applies to continuous Gaussian processes, and more specifically to the asymptotic limits $V_j$ of the empirical processes of $V_{jmn}$ defined in (3). In this section the results of Theorem 3.2 are applied to a few special examples for the purposes of conducting inference on stochastic dominance. In essence, the only ingredients needed to apply the theorem for inference is an estimate of the variance function of $V_j$, which can be made nonparametrically via a plug-in estimate, and an estimate of the contact set.

For the moment, consider just one random sample $\{X_i\}_{i=1}^n$ following distribution $F$. On compact sets the weak limit of the integrated empirical process $I_j(\cdot, \sqrt{n}(F_n - F))$ is the $j^{th}$-order integrated Brownian bridge $B^j_F := I_j(\cdot, B^j_F)$. Lemma 1 of Barrett and Donald (2003) shows that this family of processes (indexed by order $j$) have covariance functions

$$\rho_j(s, t, F) = \sum_{\ell=0}^{j-1} \binom{2j-\ell-2}{j-1} \frac{|t-s|^{\ell}}{\ell!} I_{2j-\ell-1}(s \wedge t, F) - I_j(s, F)I_j(t, F)$$

(16)

for $s, t \in T$ and 0 otherwise. Inspection of this covariance function reveals that features of these processes depend on $F$, and the laws of their suprema are affected by features of $F$ for any order $j \geq 2$. It can be verified that the covariance function $\rho_j$ defined in (16) is (at least) twice differentiable for $j \geq 2$. 

<table>
<thead>
<tr>
<th>Quantile $\tau$</th>
<th>$\tau = 0.5$</th>
<th>$\tau = 0.8$</th>
<th>$\tau = 0.9$</th>
<th>$\tau = 0.95$</th>
<th>$\tau = 0.975$</th>
<th>$\tau = 0.99$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulated</td>
<td>0.06</td>
<td>0.26</td>
<td>0.38</td>
<td>0.48</td>
<td>0.57</td>
<td>0.68</td>
</tr>
<tr>
<td>Approximated</td>
<td>0</td>
<td>0.243</td>
<td>0.37</td>
<td>0.475</td>
<td>0.566</td>
<td>0.672</td>
</tr>
</tbody>
</table>

Table 1: Quantiles of the distribution of $\sup_{t \in [0,1]} \int_0^t B_s ds$, for $B$ a standard Brownian bridge on the unit interval, as simulated in Schmid and Trede (1998), compared with the asymptotic approximation of Corollary 3.4.
For reference, the variance function is defined by

$$\sigma_j^2(t, F) := \rho_j(t, t, F) = \left(\frac{2j - 2}{j - 1}\right) I_{2j-1}(t, F) - I_j^2(t, F). \quad (17)$$

Assume now the contact set $S$ coincides with the entire range under consideration, $T$; estimated contact sets will be considered below. When no parameters are estimated, the empirical process $V_{jmn}$ actually used for testing depends on one process from each sample; Barrett and Donald (2003) show that under a few regularity conditions the process (3) converges to a process $V_j = \sqrt{\lambda}B_j^1 + \sqrt{1-\lambda}B_j^2$ (recall that $\lambda$ is the asymptotic proportion of observations from sample 1). Assuming that $S \cap T = T$, Theorem 3.2 applies to $\hat{B}^j$, implying

$$P\left\{ \sup_{t \in T} V_j(t) > u \right\} \approx\Psi\left(\frac{u}{\sigma(T)}\right) \quad (18)$$

with $\sigma^2(t) = \lambda \sigma_j^2(T,F_2) + (1-\lambda)\sigma_j^2(T,F_1)$, where the expression for the variance function from each sample, $\sigma_j^2(T,F_k)$, is defined in (17). This variance (a linear function of the distribution functions from each sample) can be estimated using plug-in estimates $F_{1m}$ and $F_{2n}$ in the place of $F_1$ and $F_2$. Since one in effect assumes the least-favorable distribution $F_1 \equiv F_2$ on $S$, it is also possible to use a pooled variance estimate to increase the power of tests, since under that hypothesis the process has a variance function identical to (17). The above formula can be used as-is to provide asymptotic $p$-values directly or inverted to find asymptotic critical values for tests and uniform confidence bands based on the supremum metric.

When there is dependence between samples, weak regularity conditions ensuring that the dependence is limited ensure that $V_j$ is a [nondegenerate] integrated Gaussian process. For example, in Linton et al. (2005) simple stationarity and mixing conditions are given, and the result (that is, for dependent samples without parameter estimation) is identical to the above display for independent samples, with only the variance term altered to $\bar{\sigma}^2(t) = \lambda \sigma_j^2(T,F_2) + (1-\lambda)\sigma_j^2(T,F_1) - 2\sqrt{\lambda(1-\lambda)}\text{Cov}\left(B_{F_1}(T), B_{F_2}(\bar{T})\right)$; this last term can be consistently estimated using the plug-in estimate

$$\hat{\text{Cov}}\left(B_{F_1}(T), B_{F_2}(\bar{T})\right) = \frac{1}{nm} \sum_k \sum_l \frac{(\bar{T} - X_k)^{j-1}(\bar{T} - Y_l)^{j-1}}{(j-1)!} I(X_k \leq \bar{T}, Y_l \leq \bar{T}) - I_j(\bar{T}, F_{1m})I_j(\bar{T}, F_{2n}). \quad (19)$$

When samples are from conditional distributions — residuals from conditional models, for example — then the finite-sample distribution of the empirical process is affected by parameter estimation and this is transmitted to the limiting distribution as well. Linton et al. (2010) provide general conditions under which the convergence of the empirical process to an integrated Gaussian process, and under those conditions the covariance function needs to be altered to specific parameter estimators, but otherwise the result is the same as the cases above. For example, assume that there are only finite-dimensional estimate that are asymptotically linear, i.e. there is a well-behaved $\psi_k$ such that

$$\sqrt{n}(\hat{\theta}_k - \theta_k) = \frac{1}{\sqrt{n}} \sum_l \psi_k(X_l, \theta) + o_p (1) \quad (20)$$
for each $k = 1, 2$, and call the . Next, assume that for each sample $k = 1, 2$, the gradient function $\hat{\Delta}_k(t, \theta_k) := \nabla \ell_j(t, \mathcal{F}_{kn}(\cdot, \theta_k))$ exists (in a left-hand sense, for the boundary of the set $\mathcal{T}$) and converges weakly to a nice limit $\Delta_k(t, \theta_k)$. When there is no dependence between samples and samples are residuals from linear regression models, Theorem 1 of Linton et al. (2005) shows that (still assuming $S = \mathcal{T}$), the formula used in the above two examples also applies with variance function $\delta^2(t)$ defined as

$$
\delta^2(t) = \lambda \left( \sigma^2_j(t, F_2) + 2E[\psi_2^\top(\theta_2)\Delta_2] + E[\psi_2^\top \Delta_2 \Delta_2^\top \psi_2] \right) + (1 - \lambda) \left( \sigma^2_j(t, F_1) + 2E[\psi_1^\top \Delta_1] + E[\psi_1^\top \Delta_1 \Delta_1^\top \psi_1] \right),
$$

(21)

where $\phi_k$ and $\Delta_k$ functions are evaluated at $t$ and $\theta_k$. Note that even though this variance function depends on the type of estimator used through the specific form of $\psi_k$ and $\Delta_k$, and the specific values of $\theta_k$, it is still possible to use Theorem 3.2 because this $V_j$ is also an integrated Gaussian process. The weakness of the conditions under which the Theorem may be applied — alternatively, the degree to which fluctuations are smoothed by integration — makes it possible to apply the same theorem in all these situations, given a plug-in estimate of the variance function evaluated at one point.

Assuming that $S \subseteq \mathcal{T}$ [strictly] makes the situation more complicated. In this case the result of Linton et al. (2010) implies that an estimate of the contact set is required. However, under the assumptions above — specifically those that ensure stochastic equicontinuity of the sequence of empirical processes — all that is needed is a consistent estimate of the upper endpoint of $S$, say $\hat{T}$ (call its limit $T_\infty$). This implies

$$
\lim_{n,m \to \infty} P \left\{ \sup_{t \in [\hat{T}, \tilde{T}]} V_{jmn}(t) > u \right\} = P \left\{ \sup_{t \in [T, T_\infty]} V_j(t) > u \right\},
$$

(22)

and that $\lim_{n,m \to \infty} \text{Var} \left( V_{jmn}(\hat{T}) \right) = \text{Var} \left( V_j(T_\infty) \right)$. Then Theorem 3.2 implies the following Corollary.

**Corollary 4.1.** Let $S$ be the contact set and suppose that $S \subseteq \text{supp}(F_1) \cup \text{supp}(F_2)$. Let $\hat{T} = T_{mn}$ be a random variable such that $\hat{T} \overset{p}{\to} T_\infty := \sup S$ and assume the regularity conditions of Linton et al. (2010). Then

$$
\lim_{u \to \infty} \lim_{n,m \to \infty} \frac{P \left\{ \sup_{t \in [\hat{T}, \tilde{T}]} V_{jmn}(t) > u \right\}}{\Psi(u/\sigma_{T_\infty})} = 1,
$$

(23)

where $\sigma^2_{T_\infty} = \text{Var} \left( V_{jmn}(T_\infty) \right)$.

This applies directly to situations in which a poverty line (contained in the contact set) is estimated from the data and used as the upper endpoint of consideration. It also implies that an estimate of the upper endpoint of the contact set is required when testing. In theory, not much is required of the contact set estimator (although in practice this can be difficult; see Section 5). Corollary 4.1 suggests an asymptotic approach to inference that has asymptotically (in level and as $m,n \to \infty$) correct size and is consistent; it is outlined below.
1. Let \( \{c_n\} \) satisfy the conditions \( c_n \to 0 \) and \( c_n \times \sqrt{n} \to \infty \), and estimate the contact set by \( S_n = \{t \in \mathcal{T} : |V_{jmn}(t)| \leq c_n\} \). Then estimate the rightmost endpoint of the contact set, \( \hat{T} = \sup S_n \).

2. Calculate \( \hat{M} = \sup_{t \in \mathcal{T}} V_{jmn}(t) \).

3. Use \( \hat{M} \) for inference via the asymptotic approximation

\[
\mathbb{P}\{\hat{M} > u\} \approx \Psi\left(\frac{u}{\hat{\sigma}(\hat{T})}\right),
\]

where \( \hat{\sigma} \) is a plug-in variance function estimator that reflects the assumptions made on the distributions (such as those discussed earlier in this section). On the contact set, it is possible to used a variance estimator that pools both samples, since it is assumed that on \( S, F_1 \equiv F_2 \).

It is important to note that \( \hat{T} \) denotes an estimate of the right endpoint of the contact set \( S \), while \( \hat{M} \) is evaluated over all of \( \mathcal{T} \). The logic of the constant sequence \( \{c_n\} \) is that when scaled by \( \sqrt{m} \), \( \{c_n\} \) tend to infinity. Because \( M := \sup_{t \in S} V_f(t) = O_p(1) \), while for \( t \in \mathcal{T} \setminus S \), under the null hypothesis \( V_{jmn}(t) \to -\infty \) and under the alternative, \( V_{jmn}(t) \to \infty \), the conditions on \( \{c_n\} \) allow one to pick out the contact set asymptotically and use it for inference. Practically speaking, only ad-hoc sequences \( \{c_n\} \) have been tried thus far and tests generally perform better with the help of a restricted domain; see Section 5 for details.

4.1 First-order stochastic dominance

This subsection illustrates the dramatic difference between the above tests and the first-order situation and also the limits to using the large deviations results for tests of first-order stochastic dominance. Because parameter values and estimates make general asymptotic results intractable, attention is restricted here to samples not affected by parameter estimates, and we simplify to the least-favorable case \( F_1 \equiv F_2 \), and refer to this common distribution simply as \( F \). This means that the limiting process in question is the time-changed Brownian bridge \( B_F \), that is, the mean-zero Gaussian process on \( \text{supp}(F) \) with covariance function \( \text{Cov}(B_F(s), B_F(t)) = F(s \wedge t) - F(s)F(t) \). Classical results on first-order dominance apply to this process over the entire support of \( F \). For restricted domains \( \mathcal{T} \), the distribution of the supremum depends on the endpoints of the interval, because \( B_F \) reaches its point of maximum variance at the median of \( F \); call this point \( m_F \). If both endpoints of \( \mathcal{T} \) are on one side of \( m_F \) (as would presumably be the case for applications that focus on the lower tail of income distributions, for example), then the point of maximal variance will be the endpoint of \( \mathcal{T} \) that is closest to the median. In the case that \( m_F \in \mathcal{T} \), the maximal variance is of course achieved there.

**Theorem 4.2.** Let \( m_F \) denote the median of the distribution \( F \). When \( m_F \in \mathcal{T} \), assume that the density \( f \) is differentiable in a neighborhood of \( m_F \). Then

\[
\lim_{u \to \infty} \frac{\mathbb{P}\{\sup_{t \in \mathcal{T}} B_F(t) > u\}}{e^{-2u^2}} = 1.
\]
When \( m_F \notin \mathcal{T} \), let \( T^* \) be the point in \( \mathcal{T} \) closest to \( m_F \). Then

\[
\lim_{u \to \infty} \frac{1}{2-2F(T^*)} \sup_{t \in \mathcal{T}} B_F(t) > u \leq \frac{\Psi(u/\sqrt{F(T^*)(1-F(T^*))})}{1-2F(T^*)} = 1. \tag{26}
\]

The first part of Theorem 4.2 is nearly the same as the textbook results based on the reflection principle (see, for example, Proposition 12.3.3 of Dudley (2002)). The major difference between the classical theorem and the above result is that the interval \( \mathcal{T} \) may be restricted to a strict subset of the support of the random variable, and for this reason, it is not exact for every level \( u > 0 \) — reflection principle arguments rely on the probability that the process crosses a level, implicitly assuming that at the left endpoint of the interval under consideration the process is almost surely zero. In the above Theorem, the argument is different in that it relies only on the property that the point of a process that contributes the overwhelming amount to the behavior of the supremum as \( u \to \infty \) is that point where the variance of the process is maximized. The second part of Theorem 4.2 is different from the first part due to the relatively equal local strength of the standard deviation and correlation functions, which have Taylor expansions of the same order when \( m_F \notin \mathcal{T} \).

5 Simulation study

In this section a few designs are investigated in a simulation study that mimics those used in Linton et al. (2005) and Linton et al. (2010); besides the examples using the processes considered in Section 3, the simulations here show that the asymptotics can be used to provide reasonably accurate and quick inference that relies on only one variance estimate.

Two parametric settings are examined for the size and power of tests and compared to existing results in the literature\(^3\). Second-order tests and third-order tests were computed using two sets of parametric designs, with parameters set equal in each distribution to check size, or set differently for a check of power. All tests have nominal 5% rejection rate, and 1000 simulation repetitions were used. The empirical processes were evaluated at each point in the combined samples; it may be possible to pick a better selection of points or calculate which midpoints between observations should also be examined, but such considerations were ignored here.

Two different sorts of tests were conducted for each design. A test over the entire observed domain of the variables was conducted, and one over the restricted range of the first and fourth deciles of the combined samples, as a simulation of the region that might typically be relevant for researchers in poverty comparisons. As will be seen below, the latter tests usually performed better than the former. This appears to be due to a complex interplay between the estimated contact set and the right tail of the empirical process, which tends to take very large values. Two distributional families are used for examples, the lognormal family and the Burr Type XII family.

\(^3\)I intend to reproduce the resampling results of other authors myself in a future version of this section
5.1 The lognormal distribution

The lognormal distribution is used often in income distribution and financial contexts, and was used in simulation experiments in Barrett and Donald (2003) and Linton et al. (2005). Here, lognormal is taken to mean that the expression $X \sim \mathcal{LN}(\mu, \sigma)$ is equivalent to the characterization $X = \exp(Y)$, where $Y \sim \mathcal{N}(\mu, \sigma^2)$. Two designs were used from the aforementioned articles, and they are summarized in Table 2 (they are designs 2a and 2d in the latter article).

<table>
<thead>
<tr>
<th></th>
<th>$X_1$</th>
<th>$X_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Design 1</strong></td>
<td>$\mathcal{LN}(0.85, 0.6^2)$</td>
<td>$\mathcal{LN}(0.85, 0.6^2)$</td>
</tr>
<tr>
<td><strong>Design 2</strong></td>
<td>$\mathcal{LN}(0.85, 0.6^2)$</td>
<td>$\mathcal{LN}(0.2, 0.1^2)$</td>
</tr>
</tbody>
</table>

Table 2: Description of lognormal designs.

The results of a simulation experiment using the first lognormal design are presented in Table 3. An investigation was made into the advantage of pooling samples for the variance estimate, and it did not appear to affect tests in any appreciable manner. Therefore only tests using variance estimates that do not pool samples are shown below.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Unrest.</th>
<th>LMW:sub</th>
<th>LMW:boot</th>
<th>Rest.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Second order</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>7</td>
<td>6.8</td>
<td>5.6</td>
<td>5.2</td>
</tr>
<tr>
<td>500</td>
<td>6.4</td>
<td>5.6</td>
<td>4.6</td>
<td>4.7</td>
</tr>
<tr>
<td>1000</td>
<td>5.3</td>
<td>6.2</td>
<td>6.5</td>
<td>4.7</td>
</tr>
<tr>
<td><strong>Third order</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>6.7</td>
<td>-</td>
<td>-</td>
<td>6</td>
</tr>
<tr>
<td>500</td>
<td>6.2</td>
<td>-</td>
<td>-</td>
<td>4.6</td>
</tr>
<tr>
<td>1000</td>
<td>4.9</td>
<td>-</td>
<td>-</td>
<td>4.8</td>
</tr>
</tbody>
</table>

Table 3: Empirical rejection probabilities of second- and third-order tests based on asymptotic approximation using the first lognormal design (where the distributions are identical). The two leftmost columns should be compared with corresponding size results from Linton et al. (2005), while the results for the restricted test and third-order test are new. Tests have a nominal 5% rate of rejection. 1000 simulation repetitions.

Table 3 reveals that the asymptotics work about as well for the least-favorable form of the null as do resampling schemes, when using the entire observed range of the data. On the other hand, the restricted test has relatively good size. A p-value plot and p-value discrepancy plot is shown for the p-values used to produce the sizes in Table 3 and are shown in Figure 1.

The p-value plots show the effect of the asymptotics in level, similarly to the integrated Brownian bridge case illustrated in Table 1. Namely, the p-values are nowhere near uniform for values that correspond to lower levels of the distribution (larger p-values) but become more exact for larger values (smaller p-values).
Figure 1: A p-value plot (left) and p-value discrepancy plot (right) for the asymptotic inference procedure. The unrestricted p-values stop at 50% because processes all begin at zero when starting from the leftmost point in the support of the random variables. Straight dotted lines are Kolmogorov-Smirnov 95th percentiles for sample size 1000, the largest one considered.

As mentioned in the previous section, a sequence of constants must be defined that converges to zero but such that when scaled to match the scaling of the empirical cdfs (by $\sqrt{m}$ or $\sqrt{n}$) the sequence diverges. The ad-hoc rule used here was

$$c_n = 10 \times \sigma(x_{\text{max}}) \times (m \wedge n)^{1/3},$$

where $x_{\text{max}}$ is the largest element of the combined samples and $\sigma(\cdot)$ is an estimate of the standard deviation function of the process. This sequence diverges more quickly than that used in Linton et al. (2010), but the sequence used in that article did not work well in this setting. The process $V_{jmn}$ can become quite wild at the largest sample observations (under the null hypothesis), and the very large, rightmost part of the integrated empirical process is so large that it is estimated to be outside of the contact set, the variance of the maximum is underestimated, and all of this leads to overrejection. The fact that the contact set was estimated too “tightly” using the other sequence could be an indication that the simulation results in Linton et al. (2010) rely on the [convex hull of the] entire support of the observed data, which is not a problem for resampling as long as the largest values do not escape too quickly (an issue demonstrated in Horváth et al. (2006)).

A picture depicting the relationship between the distributions in the second lognormal design (used for power comparison with results in Linton et al. (2010)) is shown in Figure 2. It can be seen that for the second design, neither distribution dominates the other over the whole region at either order (so the null of dominance should be rejected for unrestricted-domain tests of either order). The vertical lines in the figure represent the first and fourth deciles of the combined samples, as a simulation of the region...
that might typically be relevant for researchers in poverty comparisons. For this restricted domain, the second-order test should reject the null hypothesis because the integrated distribution functions cross in the region. However, the second distribution dominates the first at the third order, and the null should not be rejected, but the null of dominance should be rejected if the place of the distributions is reversed — that is, if one tests restricted third-order dominance of distribution 1 over distribution 2. Therefore an extra experiment was included that reverses the relationship of the distributions.

![Figure 2: The lognormal distributions in the second configuration, where they are not identical. The vertical lines represent the first and fourth deciles of the combined data, to be used for restricted testing. Both (i.e., at order 2 and 3) unrestricted tests should reject the null of dominance, while the second-order restricted test should reject and not reject at the third order (in fact, $F_1$ dominates $F_2$ at third order for this restricted domain).](image)

The results of the second lognormal design are shown in Table 4.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Unrest.</th>
<th>LMW.sub</th>
<th>LMW.boot</th>
<th>Rest.</th>
<th>Rest.rev.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Second order</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>16.8</td>
<td>2.7</td>
<td>35.4</td>
<td>-</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>91</td>
<td>33.2</td>
<td>92.4</td>
<td>-</td>
</tr>
<tr>
<td>1000</td>
<td>100</td>
<td>99.9</td>
<td>86</td>
<td>98.8</td>
<td>-</td>
</tr>
<tr>
<td><strong>Third order</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>-</td>
<td>-</td>
<td>5.6</td>
<td>12.5</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>99.6</td>
</tr>
<tr>
<td>1000</td>
<td>100</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 4: Empirical rejection probabilities of second- and third-order tests using the second lognormal design. Tests have a nominal 5% rate of rejection. 1000 simulation repetitions.

Table 4 shows empirical rejection probabilities for both tests using the second lognormal design, and
the result of tests when the distributions are reversed (in which case the test should reject the null of dominance for the reversed Design 2). There are no published results for resampling tests with which to compare the third-order results, although the similarity of the resampling and asymptotic results at second order makes it easy to presume that they look similar to the second-order comparisons reported here.

5.2 The Burr Type XII distribution

The Burr Type XII distribution has a distribution function that can be used as a stand-in for income distributions (Linton et al., 2005); it has cumulative distribution function \( F(x, \alpha, \beta) = 1 - (1 + x^\alpha)^{-\beta} \). One easy way to simulate from this distribution uses its characterization as a generalized beta distribution — the results of Xu (1995) show that if \( X \sim \text{beta}(1, \beta) \), then \( (X/(1 + X))^\beta \sim F(\cdot, \alpha, \beta) \). We used two of the designs from Linton et al. (2005), one that satisfies the least-favorable null and one that violates the null, and their parameter values are given in Table 5. They correspond to the ‘1a’ and ‘1c’ designs of Linton et al. (2005) (tabulated results are actually published in Linton et al. (2008)).

| Design 1 | \( B_{X_{11}}(4.7, 0.55) \) | \( B_{X_{11}}(4.7, 0.55) \) |
| Design 2 | \( B_{X_{11}}(4.7, 0.55) \) | \( B_{X_{11}}(2.0, 0.65) \) |

Table 5: Description of Burr Type XII designs.

Once again, Design 1 is used to check the size of the asymptotic tests, and Design 2 for power. The results of the Design 1 comparison are shown in Table 6. The same restricted domain is used for this experiment as for the previous, namely the region between the first and fourth deciles of the even mixture of both distributions.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Unrest. subsample</th>
<th>bootstrap</th>
<th>Rest.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Second order</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>4.9</td>
<td>10.1</td>
<td>6.6</td>
</tr>
<tr>
<td>500</td>
<td>6.2</td>
<td>4.9</td>
<td>5.5</td>
</tr>
<tr>
<td>1000</td>
<td>5.9</td>
<td>5.4</td>
<td>5</td>
</tr>
<tr>
<td><strong>Third order</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>500</td>
<td>5.6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1000</td>
<td>5.8</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 6: Empirical rejection probabilities of second- and third-order tests based on asymptotic approximation using the first Burr XII design (where the distributions are identical). The two leftmost columns should be compared with corresponding size/power results from Linton et al. (2005), while the results for the restricted test are new. Tests have a nominal 5% rate of rejection. 1000 simulation repetitions.

Table 6 reveals that for this design the asymptotic tests are a bit conservative for restricted tests and
larger sample sizes, but still marginally satisfactory. The third-order results look as well-behaved as the second-order. Another pair of p-value plots is shown in Figure 3.

Figure 3: A p-value plot and p-value discrepancy plot over the lower fifth of the theoretical p-value distribution for the first Burr XII design. Dotted lines are Kolmogorov-Smirnov bands using $n = 1000$, the largest sample size that was considered.

In Figure 3 the effect of the asymptotic approximation in levels can be seen again, and the results are similar to the first lognormal design. The behavior of the restricted tests is a little less close to the theoretical (uniform) distribution function of the p-values, compared to the lognormal case, but still well within the Kolmogorov-Smirnov bands for the largest sample size.

Design 2 is used to compare test procedures for power. A plot of the theoretical integrated distribution functions and their difference is shown in Figure 4. It shows that all tests should reject the null of dominance of the second distribution over the first, although in the region in which the null is violated, the degree of violation is relatively small compared to the larger differences in the tails (the figures are truncated on the right so that the region where the null is violated is visible).

The region where the null is violated in the third-order figure the difference is quite small in magnitude, relative to the much larger negative differences to the right. This should make the unrestricted third-order test rather challenging, while for the chosen restricted range the third-order test should have fairly good power against this alternative. The results of the second Burr XII design are shown in Table 7.

As could be predicted from reflecting on Figure 4, the unrestricted-domain third-order test has no power against this alternative. On the other hand, the same argument could be made by looking at the second-order analog, and the power against the alternative at that order is quite good. It is probable that a better estimator of the contact set is required for tests of higher order (the ad-hoc rule used was fine-tuned using second-order tests). The asymptotic approximation does at least as well as the resampling schemes considered in Linton et al. (2010), although the comparison is not quite exact because the
Figure 4: The Burr Type XII distributions in the second configuration, where they are not identical. The vertical lines represent the first and fourth deciles of the combined data, to be used for restricted testing. On the unrestricted domain, neither (i.e., at order 2 and 3) test should reject the null of dominance of $F_2$ over $F_1$. Therefore for power considerations the null of dominance of $F_1$ over $F_2$ is posited. Meanwhile, both tests should reject the null of dominance of $F_2$ over $F_1$ on the restricted domain.

The contact set estimator uses a different rule than in that article.

The plan for a later draft of this paper is to include replications of estimators based resampling schemes used by other authors (including the multiplier bootstrap of Barrett and Donald (2003)), one more simulation design and an example using data from the Luxembourg Income Study, which will hopefully reinforce the message here, which is that the asymptotic methods proposed above are reasonably accurate in many situations.

6 Conclusion

Large deviations asymptotic methods that apply to smooth Gaussian processes can be used to conduct asymptotic inference for stochastic dominance orderings of all levels higher than the first order. The methods suggested by the theory are quite simple, and the simulation study of Section 5 provides some evidence that they are as accurate as previously-proposed resampling schemes suggested in the literature. Resampling methods can and should be used when asymptotics are unknown or deemed unreliable, although the simulation study of Section 5 provides limited evidence that the asymptotics can reasonably provide a first-pass test of stochastic dominance.
<table>
<thead>
<tr>
<th>Sample size</th>
<th>Unrest.</th>
<th>subsample</th>
<th>bootstrap</th>
<th>Rest.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Second order</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>16.8</td>
<td>2.7</td>
<td>35.4</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>91</td>
<td>33.2</td>
<td>92.4</td>
</tr>
<tr>
<td>1000</td>
<td>100</td>
<td>99.9</td>
<td>86</td>
<td>98.8</td>
</tr>
<tr>
<td><strong>Third order</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>11</td>
</tr>
<tr>
<td>500</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>63.9</td>
</tr>
<tr>
<td>1000</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>97.6</td>
</tr>
</tbody>
</table>

Table 7: Empirical rejection probabilities of second- and third-order tests using the second Burr XII design. The third-order test has no power against this alternative, probably due to imprecise estimation of the contact set. On the other hand, restricted tests do quite well. Tests have a nominal 5% rate of rejection. 1000 simulation repetitions.

A Proof of results in the text

Proof of Lemma 3.1. Let \( \rho \) be the covariance function of \( \xi \) and \( \psi \) be the covariance function of \( \xi' \). By definition, \( \xi' \) is the pathwise derivative of \( \xi \), and this implies that (via (5))

\[
\rho(s, t) = \int_0^s \int_0^t \psi(x, y) \, dx \, dy. \tag{28}
\]

Consider the variance of the integrated process, \( \sigma^2_\xi(t) \equiv \rho(t, t) \). Taking a derivative of this function one finds

\[
\frac{d}{dt} \rho(t, t) = \frac{\partial}{\partial s} \rho(s, t)|_{s=t} + \frac{\partial}{\partial t} \rho(s, t)|_{s=t} = 2 \int_0^t \psi(x, t) dx,
\tag{29}
\]

using the symmetry of \( \psi \) and (28). The above integral can be approximated by sums using a sequence of partitions such that the mesh diameter shrinks to zero. Choose \( 0 = x_0 \leq x_1 \leq \ldots \leq x_n = T \) and let \( \Delta_i := x_i - x_{i-1} \) for \( i = 1, \ldots n \). Then define

\[
S_n = \sum_{i=1}^n \psi(x_i, t) \Delta_i. \tag{30}
\]

Because \( \psi \) is a covariance function, it must be a “nonnegative-definite type” (Loéve, 1978, §37; the name arising because nonnegative definite functions only rely on one argument) — that is, the matrix \( P \in \mathbb{R}^{n \times n} \) defined by \( \{P_{ij}\} = \{\psi(t_i, t_j)\} \), for any collection \( t \in \mathbb{R}^n \), is positive definite. This means also that \( P \) is a diagonally dominant matrix: for any row \( i \), \( |P_{ii}| \geq \sum_{k \neq i} |P_{ik}| \) and analogously for columns of \( P \). This implies \( S_n \geq 0 \) for each \( n \), and since \( S_n \to d/dt \sigma^2_\xi(t) \) as \( n \to \infty \), the derivative must be monotone increasing because \( \psi \) is continuous and \( T \) is compact.

Proof Theorem 3.2. Because the process is the integral of another process, it is mean-square differentiable and pathwise differentiable, and its variance is monotonically increasing over its domain, so that
$ar{T}$ is the point of maximal variance. The Theorem of Talagrand (1988) shows that the result of this
Theorem is equivalent to the condition that

$$
\lim_{\delta \to 0} \frac{1}{\delta} \mathbb{E} \left[ \sup_{t \in T_\delta} \xi_t - \xi_{\bar{T}} \right] = (31)
$$

where

$$
T_\delta = \{ t \in \mathcal{T} : \rho(t, \bar{T}) \geq \rho(\bar{T}, \bar{T}) - \delta^2 \}. \quad (32)
$$

Taking a two-term Taylor expansion of $\rho$ in its first argument (the second-order expansion will be
needed in one of two cases below),

$$
\rho(t, \bar{T}) = \rho(\bar{T}, \bar{T}) - \rho_{10}(\bar{T}, \bar{T})(\bar{T} - t) + \frac{1}{2} \rho_{20}(\bar{T}, \bar{T})(\bar{T} - t)^2 (1 + o(1)), \quad t \nearrow \bar{T}. \quad (33)
$$

In case $\text{Var} \left( \xi'_{\bar{T}} \right) > 0$, the one-term expansion is sufficient, since then $\rho_{10} = \text{Cov} \left( \xi'_{\bar{T}}, \xi_{\bar{T}} \right) > 0$. Therefore for some $\delta$ small enough, there is a $K > 0$ such that

$$
\delta^2 \geq \rho(\bar{T}, \bar{T}) - \rho(t, \bar{T}) \approx \rho_{10}(\bar{T}, \bar{T})(\bar{T} - t) \Rightarrow t \geq \bar{T} - K\delta^2. \quad (34)
$$

Therefore

$$
\lim_{\delta \to 0} \frac{1}{\delta} \mathbb{E} \left[ \sup_{t \in T_\delta} \xi_t - \xi_{\bar{T}} \right] = \lim_{\delta \to 0} \frac{1}{\delta} \mathbb{E} \left[ \sup_{t \in T_\delta} \int_t^{\bar{T}} \xi'_s ds \right] \leq \lim_{\delta \to 0} \frac{1}{\delta} \mathbb{E} \left[ \sup_{t \in T_\delta} \int_t^{\bar{T}} |\xi'_s| ds \right] \leq \lim_{\delta \to 0} \frac{1}{\delta} \mathbb{E} \left[ \int_{\bar{T} - K\delta^2}^{\bar{T}} |\xi'_s| ds \right]. \quad (35)
$$

(36)

(37)

However, because the process is bounded, $\sup_{t \in \mathcal{T}} \mathbb{E} \left[ \xi'_t \right] < \infty$, so the above integral is $O(\delta^2)$ and the limit in question is 0.

In the case that $\lim_{t \nearrow \bar{T}} \text{Var} \left( \xi'_t \right) = 0$, first note that this implies $\xi'_t = o_P(1)$ as $t \to \bar{T}$ since the
derivative process is assumed to be mean-zero. The calculations based on a one-term Taylor expansion
do not apply because this condition implies $\rho_{10}(\bar{T}, \bar{T}) = \text{Cov} \left( \xi'_{\bar{T}}, \xi_{\bar{T}} \right) = 0$. Instead using the full two-
term expansion shown above, analogous calculations show that for some $K > 0$ (not necessarily the
same as the previous $K$),

$$
t \in T_\delta \Rightarrow t \geq \bar{T} - K\delta. \quad (38)
$$
Then similar calculations also imply

$$\lim_{\delta \to 0} \frac{1}{\delta} E \left[ \sup_{t \in T} \xi_t - \xi_T \right] \leq \lim_{\delta \to 0} \frac{1}{\delta} E \left[ \int_{T-k\delta}^{T} |\xi'_s| ds \right].$$

(39)

In this case, however, \( \xi'_T = o_p(1) \), and

$$\int_{T-k\delta}^{T} |\xi'_s| ds = o_p(\delta),$$

(40)

implying the expectation of this integral is \( o(\delta) \) and further that the same result holds for this case as for the above case when it was assumed that the variance of \( \xi' \) was nonzero at \( T \). \( \blacksquare \)

**Proof of Theorem 4.2.** The correlation function of \( B_F \) has the following expansion, as \( s \to t \), for all \( t \in T \):

$$\text{Corr}(B_F(s), B_F(t)) = 1 - \frac{f(t)}{2\sigma^2(t, F)} |s-t|(1 + o(1)), \quad s \to t.$$ 

The standard deviation function has the following second-order expansion (which is needed in one of two cases below):

$$\sigma_1(t, F) = \sigma_1(T^*, F) + \frac{1}{2\sigma_1(T^*, F)} \left[ f(T^*)(1 - 2F(T^*)) (t - T^*) \right.$$

$$+ \frac{1}{2\sigma(T^*, F)} \left[ \frac{-1}{4\sigma^2(T^*, F)} f(T^*)(1 - 2F(T^*)) + \frac{1}{2} f'(T^*)(1 - 2F(T^*)) - f^2(T^*) \right] (t - T^*)^2(1+o(1)), \quad t \to T^*.$$ 

(41)

When \( m_F \in \mathcal{T} \), the fact that as the median solves \( F(m_F) = 1/2 \) implies that the first term in the above expansion is zero, and the variance function has a second-order expansion (for which it is necessary to assume that the density function is differentiable in a neighborhood of \( m_F \)). At all other points a first-order expansion is sufficient (which means no differentiability is required for \( f \)).

The order of the variance function expansion is 2, which is greater than the order of the correlation function expansion. This means that Theorem 8.2 (i) applies to this function. That is, when \( m_F \in \mathcal{T} \),

$$\mathbb{P} \left\{ \sup_{t \in \mathcal{T}} B_F(t) > u \right\} = H_m \frac{u}{\sigma^2(m_F, F)} \Psi \left( \frac{u}{\sigma^2(m_F, F)} \right) (1 + o(1)), \quad u \to \infty,$$

(42)

where

$$H = \int_{\mathbb{R}} \exp \left\{ -2\sigma_1^2(m_F, F)t^2 \right\} dt.$$ 

(43)

Recalling that the variance at \( m_F \) is 1/4, all of the above expressions can be rewritten to imply the first result.

When \( m_F \not\in \mathcal{T} \), the situation is different. The point of maximal variance is \( T^* \) and the variance
function has a non-negligible first-order expansion at $T^*$, the same order as the correlation function expansion. Using Theorem 8.2 (ii) of Piterbarg (1996), one finds that

$$P \left\{ \sup_{t \in T} B(t) > u \right\} = H' \Psi \left( \frac{u}{\sigma_1(T^*, F)} \right) (1 + o(1)), \quad u \to \infty,$$

(44)

where $H'$, in terms of the notation in Piterbarg (1996), is equal to $H_1^{1-2F(T^*)}((-\infty, 0]) = H_1^{1-2F(T^*)}([0, \infty))$, and

$$H' = E \left[ \exp \left\{ \max_{(0, \infty)} \left( \chi - (1 - 2F(T^*))t \right) \right\} \right],$$

(45)

and $\chi$ is a Brownian motion with mean function $E[\chi(t)] = -|t|$ and covariance $Cov(\chi(s), \chi(t)) = |t| + |s| - |t - s|$. $H'$ is related to "standard" Pickands constants (cf. Pickands (1969), Piterbarg (1996)) and although the value of those constants are known only in a small number of cases, the calculation of $H'$ above happens to be tractable in this case. Note that $\chi$ is a scaled Brownian motion with drift — it has the same distribution as $p_2 W - t$, where $W$ is a standard Brownian motion on the positive half line.

For any $d > 0$,

$$\lim_{T \to \infty} P \left\{ \sup_{[0, T]} \chi_t - dt \geq b \right\} = \lim_{T \to \infty} P \left\{ \max_{[0, T]} \sqrt{2} W_t - (1 + d)t \geq b \right\} = \lim_{T \to \infty} P \left\{ \max_{[0, T]} W_t - \frac{1 + d}{\sqrt{2}} t \geq b/\sqrt{2} \right\}.$$  

(46), (47)

The distribution of the supremum of drifting Brownian motion is known; see Karatzas and Shreve (1998, Section 3.5.C, exercise 5.9) or Jeanblanc et al. (2009, Chapter 3). Specifically, for standard Brownian motion with drift $\mu$, that is, $\tilde{W}_t = W_t - \mu t$ for $\mu > 0$, the density function of $\sup_{t \in [0, \infty)} \tilde{W}_t$ is exponential:

$$P \left\{ \sup_{t \in \mathbb{R}_+} \tilde{W}_t \in dB \right\} = 2\mu e^{-2\mu b} db.$$  

(48)

Then using dominated convergence this means

$$P \left\{ \sup_{t \in \mathbb{R}_+} \chi_t - dt \in dB \right\} = \sqrt{2} \frac{1 + d}{\sqrt{2}} e^{-\sqrt{2}(1+d)b/\sqrt{2}} db = (1 + d)e^{-(1+d)b} db.$$  

(49)

Then

$$E \left[ \exp \left\{ \sup_{R_+} \chi_t - |dt| \right\} \right] = (1 + d) \int_0^\infty e^x e^{-(1+d)x} dx = d + 1.$$  

(50)

Therefore when $d = 1 - 2F(T^*)$, the constant can be calculated as that in the statement of the Theorem.
References


