A contamination model for approximate stochastic order.*

Pedro C. Álvarez-Esteban[1], E. del Barrio[1], J.A. Cuesta-Albertos[2] and C. Matrán[1]

[1] *Departamento de Estadística e Investigación Operativa and IMUVA, Universidad de Valladolid*

[2] *Departamento de Matemáticas, Estadística y Computación, Universidad de Cantabria*

April 21, 2014

Abstract

Stochastic ordering among distributions has been considered in a variety of scenarios. Economic studies often involve research about the ordering of investment strategies or social welfare. However, as noted in the literature, stochastic orderings are often a too strong assumption which is not supported by the data even in cases in which the researcher tends to believe that a certain variable is somehow smaller than other. Instead of considering this rigid model of stochastic order we propose to look at a more flexible version in which two distributions are said to satisfy an approximate stochastic order relation if they are slightly contaminated versions of distributions which do satisfy the stochastic ordering. The minimal level of contamination that makes this approximate model hold can be used as a measure of the deviation of the original distributions from the exact stochastic order model. Our approach is based on the use of trimmings of probability measures. We discuss the connection between them and the approximate stochastic order model and provide theoretical support for its use in data analysis, including asymptotic distributional theory as well as non-asymptotic bounds for the error probabilities of our tests. We also provide simulation results and a case study for illustration.

*Research partially supported by the Spanish Ministerio de Ciencia y Tecnología, grants MTM2011-28657-C02-01 and MTM2011-28657-C02-02, and by Consejería de Educación de la Junta de Castilla y León, grant VA212U13.

1 Introduction

Stochastic order relations between distributions have been considered in a great variety of scenarios. Clinical studies are usually related to degrees of disease linked to different behaviors that often can be ordered through suitable variables. Economic studies frequently involve order relations on variables measuring, e.g., investment strategies or social welfare. In any case, an stochastic order indicates a global relation between two distributions that improves those based on making comparisons through individual indices or features of the distributions. The stochastic order between two distributions was introduced in Lehmann (1955). In terms of distribution functions, F, G, it is defined by

$$F \leq_{st} G \quad \text{if and only if} \quad F(x) \geq G(x) \text{ for all } x \in \mathbb{R} \tag{1}$$

(the inequality would be strict if $F(x) < G(x)$ at least for any x). The books by Shaked and Shanthikumar (2007) and Müller and Stoyan (2002) provide a rather complete overview of this topic, including a discussion of a great variety of other stochastic orders. However, our starting point in this paper is that, as noted in Arcones et al (2002), these orderings are in general too strong as an assumption in problems in which one is inclined to believe that a population X is somehow smaller than another population Y. In other words, the stochastic order is a 0-1 relation, that either holds or not. We believe that some index of the level of agreement with the stochastic order model for intermediate situations can be helpful.

Let us focus, for simplicity, on the two-sample problem, where two independent samples are obtained from F and G. From a methodological point of view the statistical testing problems of interest in relation with the stochastic order are (up to minor variations)

a) H_0: $F = G$, versus H_a: $G >_{st} F$

b) H_0: $G \geq_{st} F$, versus H_a: $G \not\geq_{st} F$

c) H_0: $G \not\geq_{st} F$, versus H_a: $G \geq_{st} F$

Problem a), usually referred to as the one-sided test, assumes that an stochastic ordering holds and the focus is put on giving enough statistical evidence that the relation is strict. Sometimes such an assumption is scarcely justified, or it is merely the result of the intuition of practitioners. Even if obvious, it is relevant to say that some caution should be adopted in such cases: unlike problems b) or c), for arbitrary distribution functions F and G both H_0 and H_a can be false.

Testing for stochastic dominance is the usual way of making reference to problem b), which is the goal of a sequence of papers beginning with McFadden (1989) and including Mosler (1995), Anderson (1996), Davidson and Duclos (2000), Schmid and Trede (1996 a,b), Barrett and Donald (2003), Linton et al. (2005), Linton et al. (2010), among others. It has the same statistical meaning as a goodness of fit problem. We just look for absence of evidence against our stochastic order hypothesis as a minimal requirement to continue our analyses under such assumption. Some weaknesses of this approach are well known and lead to exploring alternative or complementary tools as we will recall below.

Problem c) appears to be the most attractive for people interested in assessing the existence of an stochastic order between the parent distributions, because rejecting the null would provide convincing evidence to guarantee that G stochastically dominates F. Unfortunately, as often happens when testing hypotheses, searching for a well behaved α-level test for this problem would be unpractical: the 'no data' test, rejecting H_0 with probability α regardless of the data is the UMP test. This is showed in Berger (1988) in the one-sample setting, but the result can be easily generalized to the two-sample setup. The workaround used there to overcome this problem was testing 'restricted stochastic dominance', that is, testing the property on a fixed closed interval excluding the tails of the distribution. A similar approach has been considered in the two-sample setting in Davidson and Duclos (2013).

Here we address the problem of assessing stochastic order between two distributions as in problems b) and c) from a new perspective based on contaminated (mixture) models. More than an alternative technique for testing the null models b) and c), our goal is to provide (through very simple techniques) additional resources to the available procedures for the analysis of stochastic dominance. More precisely, for $\pi \in (0,1)$, we consider the model

$$F = (1-\pi)\tilde{F} + \pi H, \quad \tilde{F} \leq_{st} G, \tag{2}$$

where $\tilde{F}$ and H are distribution functions, and some other suitable variations of it. If model (2) holds true (for small π) then the stochastic order model (1) holds except for a small fraction of observations coming from F and, in this sense, we could say that the stochastic order model is essentially valid. Alternatively, we could consider the minimal value of π such that (2) holds. This yields a measure of deviation from the original model (1). In this paper we provide appropriate statistical tools for analysis and inference about this model as well as for suitable versions for the two-sample case.

This approach is new in this setting although it has been already considered in statistical testing in contingency tables and multinomial parametric models (Rudas et al. (1994) and Liu and Lindsay (2009)) or in the analysis of similarity between samples in a fully nonparametric context (Álvarez-Esteban et al. (2012)). In fact it is closely related to ideas that go back to Hodges and Lehmann (1954) and their discussion of practical vs. statistical significance. We further elaborate on this idea in Section 2 below.

Our handling of contaminated models is based on dealing with the dual idea of trimmings of a probability, as introduced in Álvarez-Esteban et al. (2008). In general, sets of trimmings have nice statistical properties, see, e.g., Álvarez-Esteban et al. (2012). We will show that they also behave well in the setting of stochastic ordering.

As a matter of example we discuss on the evolution with age of the commonly assumed fact that generally boys are taller than girls. Our analysis is focused on a data set obtained from NHANES (National Health and Nutrition Examination Survey) and shows statistical evidence of the levels of agreement of data with the stochastic dominance of heights of boys over those of girls.

The organization of this paper is as follows. Section 2 provides some general background on trimmings, their connections to contaminated models and, in particular, to contaminated stochastic order models. We also discuss links to related work on approximately valid models. In Section 3 we give asymptotic theory for approximate inference and non-asymptotic bounds related to contaminated stochastic order models, although the involved proofs are postponed to a technical appendix. Finally, Section 4 contains simulation results, the analysis of NHANES heights data set and some final conclusions.

2 Stochastic dominance and trimming

Trimming procedures are one of the main tools in Robust Statistics for their adaptability to a variety of statistical problems. By trimming according to a particular pattern we downplay the influence of contaminating data in our inferences. The introduction of data-dependent versions of trimming, often called impartial trimming, allows to overcome some limitations of earlier versions of trimming which simply removed extreme observations at tails. Generally, impartial trimming is based on some optimization criterion, keeping the fraction of the sample (of a prescribed size) which yields the least possible deviation with respect to a theoretical model (see e.g.Álvarez-Esteban et al. (2008, 2012); Cuesta et al. (1997); García-Escudero et al. (2008); Maronna (2005); Rousseeuw (1985)).

Trimming a fraction π of a sample or data set of size n usually means replacing the empirical measure by a new one in which the data are re-weighted so that the trimmed points have now zero probability while the remaining points will have weight $1/n(1-\pi)$. Instead of simply keeping/removing data we can increase the weight of data in good ranges (by a factor bounded by $1/(1-\pi)$) and downplay the importance of data in bad zones, not necessarily removing them. If the random generator of the sample were P, the theoretical counterpart of the trimming procedure would be to replace the probability $P(B) = \int_B 1\, dP$ by the new probability

$$\tilde{P}(B) = \int_B g\, dP \text{ where } 0 \leq g \leq \frac{1}{1-\pi} \; P\text{-almost surely.} \tag{3}$$

We call a probability measure like $\tilde{P}$ in (3) a π-*trimmed* version or a π-trimming of P. The set of π-trimmings of P will be denoted by $\mathcal{R}_\pi(P)$. If $\pi = 0$ then $\mathcal{R}_\pi(P) = \{P\}$. If $\pi = 1$ then we keep the notation $\mathcal{R}_1(P)$ for the set of probabilities which are absolutely continuous with respect to P. This definition of trimming has been considered by several authors (see, e.g., Gordaliza (1991); Álvarez-Esteban et al. (2008)). The flexibility in allowing downweighting rather than removing points results in nice properties of $\mathcal{R}_\pi(P)$ including, in particular, a link between contaminated models and sets of trimmings as we show in the next result.

Proposition 2.1 *Let P_0, P be probability distributions on $\mathbb{R}$ with distribution functions F_0 and F, respectively and $\pi \in [0,1)$. The following statements are equivalent:*

a) $P = (1-\pi)P_0 + \pi Q$ *for some probability* Q.

b) $(1-\pi)P_0(B) \leq P(B)$ *for every measurable* B.

c) $P_0 \in \mathcal{R}_\pi(P)$.

d) $F(x) = (1-\pi)F_0(x) + \pi G(x)$ *for every* $x \in \mathbb{R}$, *for some distribution function* G.

Proof. Assume a) holds so that $P = (1-\pi)P_0 + \pi Q$. Since Q is a probability, then $P \geq (1-\pi)P_0$ holds. Under condition b) P_0 is absolutely continuous with respect to P. Hence, by the Radon-Nikodym theorem, there exists a nonnegative density function, say $g := \frac{dP_0}{dP}$, such that $P_0(B) = \int_B g\ dP$ for every B. Consider the set $B = \{g > \frac{1}{1-\pi}\}$. If $P(B) > 0$ and $B_\delta = \{g \geq \frac{\delta}{1-\pi}\}$, then, $P(B_\delta) > 0$ for some $\delta > 1$, thus

$$(1-\pi)P_0(B_\delta) = (1-\pi)\int_{B_\delta} g\ dP \geq (1-\pi)\frac{\delta}{1-\pi}P(B_\delta) > P(B_\delta),$$

which contradicts $P \geq (1-\pi)P_0$. Therefore, $g \leq \frac{1}{1-\pi}$ $P-$almost surely and c) holds.

If c) holds, let g be the density of P_0 with respect to P which, by (3), satisfies $P_0(B) = \int_B g\ dP$ and $0 \leq g \leq \frac{1}{1-\pi}$. Then $(1-\pi)P_0(B) \leq \int_B 1\ dP = P(B)$ and we can define the nonnegative measure $\tilde{Q}(B) := P(B) - (1-\pi)P_0(B)$. Now set $Q(B) := \tilde{Q}(B)/\pi$, and the decomposition a) follows. Finally note that statements a) and d) are trivially equivalent. $\square$

We note that the equivalence of a), b) and c) holds in greater generality than presented here. Since we are interested in the connection to stochastic order we refrain from pursuing this issue here. We note also that the contaminated model a) is not symmetric in P and P_0. In contrast, the consideration of similarity between two probabilities, as introduced in Álvarez-Esteban et al. (2012) is a symmetric concept. We will return to this later in this section.

Statement d) in Proposition 2.1 involves only distribution functions which are the relevant objects in assessment of stochastic order. So, for the sake of economy, we will often say 'F_0 is a trimmed version of F' or write '$F_0 \in \mathcal{R}_\pi(F)$' to mean the related fact for the associated probabilities.

Let us assume that some model distribution, F_0, say, is given. We might be interested in assessing whether the random generator of a sample, F, satisfies some stochastic order relation with respect to F_0. As noted in the Introduction, model (1) is possibly a too rigid model to be realistic and we could, instead, consider model (2), namely,

$$F = (1-\pi)G + \pi H, \quad \text{for some } G \leq_{st} F_0. \tag{4}$$

Just as in Proposition 2.1, the contaminated model (4) can be simply formulated in terms of trimming. With this goal, we write $\mathcal{F}_{st}(F_0)$ for the set of distribution functions that are stochastically smaller than F_0, that is,

$$\mathcal{F}_{st}(F_0) = \{G : G \leq_{st} F_0\}.$$

Then (4) holds if and only if

$$\mathcal{R}_\pi(F) \cap \mathcal{F}_{st}(F_0) \neq \emptyset \tag{5}$$

and this provides an alternative description of the contaminated model (4) in terms of trimmings.

So far, our contamination model deals with distributions in an asymmetric way. We take one of them, F_0, as a reference model and wonder whether the other one is, after suitable trimming, stochastically majorated by the model. However, with applications to two-sample problems in mind, we should define a notion of proximity to the stochastic dominance model on the basis of both distributions (or of both samples) without a predetermined reference model. In the two-sample similarity problem considered in Álvarez-Esteban et al. (2012), similarity of P_1 and P_2 at level π means that there exist probabilities Q, P_1', P_2' such that $P_1 = (1-\pi)Q + \pi P_1', P_2 = (1-\pi)Q + \pi P_2'$, that is, that P_1 and P_2 are π-contaminated versions of a common distribution. This suggests that in the present setup of stochastic order we consider the model

$$\begin{cases} F_1 &= (1-\pi)G_1 + \pi F_1' \\ \\ F_2 &= (1-\pi)G_2 + \pi F_2', \end{cases} \quad \text{for some } G_1, G_2 \text{ such that } G_1 \leq_{st} G_2. \tag{6}$$

This contaminated model can be described, as well, in terms of trimmings. In fact, if $\mathcal{F}_{st}$ denotes the set of pairs of distribution functions (G_1, G_2) such that $G_1 \leq_{st} G_2$, then it follows from Proposition 2.1 that (6) holds if and only if

$$(\mathcal{R}_\pi(F_1) \times \mathcal{R}_\pi(F_2)) \cap \mathcal{F}_{st} \neq \emptyset. \tag{7}$$

It is convenient at this point to assign names and notation to the contaminated stochastic order models.

Definition 2.2 *For distribution functions F_1 and F_2 we say that F_1 is stochastically smaller than F_2 at level π and write $F_1 \leq_{st}^{\pi} F_2$ if (6) (equivalently, if (7)) holds.*

Furthermore, for distribution functions F, F_0, we say that F is a π-level stochastic lower bound of F_0 and write $F \in SLB_\pi(F_0)$ if (4) or, equivalently, if (5) holds.

We note that these models can be adapted in a straighforward way to deal with contaminated stochastic minorization instead of majorization. We will use these models in the sequel with corresponding adapted notation such as $F \in SUB_\pi(F_0)$ or $F_1 \geq_{st}^{\pi} F_2$.

We note that $F_1 \geq_{st}^{\pi} F_2$ if and only if $F_2 \leq_{st}^{\pi} F_1$. A bit more caution is needed with the set $SUB\pi(F_0)$, since $F \in SUB\pi(F_0)$ and $F_0 \in SLB_\pi(F)$ are not equivalent.

We provide now simple evidence that the formulation of contaminated stochastic order in terms of trimmings is particularly convenient. While two different trimmings of a fixed probability are not necessarily comparable in stochastic order, our next result shows that the set of trimmings of a fixed probability has a minimum and a maximum for the stochastic order.

Proposition 2.3 *Consider a distribution function F and $\pi \in [0,1)$. Define the distribution functions*

$$F^\pi(x) = \begin{cases} 0 & \text{if } x < F^{-1}(\pi) \\ \frac{1}{1-\pi}(F(x)-\pi) & \text{if } x \geq F^{-1}(\pi) \end{cases}$$

and

$$F_\pi(x) = \begin{cases} 1 & \text{if } x \geq F^{-1}(1-\pi) \\ \frac{1}{1-\pi}F(x) & \text{if } x < F^{-1}(1-\pi) \end{cases},$$

where F^{-1} denotes the quantile function associated to F, namely, $F^{-1}(t) = \inf\{x: \ t \leq F(x)\}$. Then $F^\pi, F_\pi \in \mathcal{R}_\pi(F)$ and any other $\tilde{F} \in \mathcal{R}_\pi(F)$ satisfies

$$F_\pi \leq_{st} \tilde{F} \leq_{st} F^\pi.$$

Proof. Consider F_π. It is easy to see that $\frac{1}{\pi}(F-(1-\pi)F_\pi)$ is a distribution function, which, by Proposition 2.1, shows that $F_\pi \in \mathcal{R}_\pi(F)$. Any other trimming of F, say $\tilde{F}$, can be expressed (recall (3)) as $\tilde{F}(x) = \int_{-\infty}^x g(t)dF(t)$ for some density g (w.r.t. P) satisfying $0 \leq g \leq \frac{1}{1-\pi}$. But then $\tilde{F}(x) \leq \min(\frac{1}{1-\pi}F(x), 1) = F_\pi(x)$ for all x, that is, $F_\pi \leq_{st} \tilde{F}$. The claims about F^π follow similarly. $\square$

An interesting consequence of Proposition 2.3 is that one can check whether the contaminated stochastic order models hold by looking just at the extremes of the relevant sets of trimmings. This, in turn, provides very simple characterizations of the minimal contamination level required for a contaminated stochastic order model to hold. We give details about this facts in our next results. This minimal contamination level under which some stochastic order relation holds is a useful measure of the deviation from the (pure) stochastic order model and will be used in later sections.

Proposition 2.4 *For arbitrary distribution functions, F, F_0, and $\pi \in [0,1)$ the following are equivalent:*

(a) $F \in SLB_\pi(F_0)$ *(b)* $F_\pi \leq_{st} F_0$ *(c)* $\pi \geq \pi_0$,

where

$$\pi_0 := \sup_{x:\, F_0(x)>0} \frac{F_0(x)-F(x)}{F_0(x)}.$$

In particular, π_0 is the minimal value of π for which $F \in SLB_\pi(F_0)$.

Proof. If (a) holds then there exists $G \in \mathcal{R}_\pi(F_0)$ such that $G \leq_{st} F_0$. But then, by Proposition 2.3, $F_\pi \leq_{st} G \leq_{st} F_0$ and we have (b). Obviously, (b) implies (a) since $F_\pi \in \mathcal{R}_\pi(F)$. Now, (b) is equivalent to

$$\frac{1}{1-\pi} F(x) \geq F_0(x)$$

for every $x \in \mathbb{R}$ (note that the inequality holds trivially for $x \geq F^{-1}(1-\pi)$). But this is, in turn, equivalent to

$$\pi \geq \frac{F_0(x) - F(x)}{F_0(x)}$$

for all $x \in \mathbb{R}$ for which $F_0(x) > 0$. This shows that (b) and (c) are equivalent too and completes the proof. □

Remark 2.4.1 In a completely symmetric fashion we could check that $F \in SUB_\pi(F_0)$ if and only if

$$\pi \geq \pi_0' = \sup_{x:\, F_0(x)<1} \frac{F(x) - F_0(x)}{1 - F_0(x)} \tag{8}$$

so that π_0' is the minimal contamination level required for $F \in SUB_\pi(F_0)$ to hold.

We deal in the next result with the $\leq_{st}^\pi$ model. It can be proved mimicking the proof of Proposition 2.4, hence, we omit details.

Proposition 2.5 *For arbitrary distribution functions, F_1, F_2, and $\pi \in [0,1)$ the following are equivalent:*

(a) $F_1 \leq_{st}^\pi F_2$ *(b)* $(F_1)_\pi \leq_{st} (F_2)^\pi$ *(c)* $\pi \geq \pi(F_1, F_2)$, *where*

$$\pi(F_1, F_2) := \sup_{x \in \mathbb{R}} (F_2(x) - F_1(x)).$$

In particular, $\pi(F_1, F_2)$ is the minimal value of π for which $F_1 \leq_{st}^\pi F_2$.

As in Remark 2.4.1, we can check that $F_1 \geq_{st} \pi F_2$ if and only if $\pi \geq \tilde{\pi}(F_1, F_2)$ $:= \sup_{x\in\mathbb{R}}(F_1(x) - F_2(x)) = \pi(F_2, F_1)$. We note also that, since

$$\begin{aligned} \sup_{x\in\mathbb{R}} |F_1(x) - F_2(x)| &= \max\{\sup_{x\in\mathbb{R}} (F_2(x) - F_1(x)), \sup_{x\in\mathbb{R}} (F_1(x) - F_2(x))\} \\ &= \max\{\pi(F_1, F_2), \pi(F_2, F_1)\}, \end{aligned}$$

the relations $F_1 \leq_{st,\pi_1} F_2$ and $F_1 \geq_{st,\pi_2} F_2$ imply $\sup_{x\in\mathbb{R}} |F_1(x) - F_2(x)| \leq \max\{\pi_1, \pi_2\}$. Hence, if π_1 and π_2 are small, then F_1 and F_2 are close to each other (in Kolmogorov distance).

Next, we provide some examples to illustrate the meaning of the contaminated stochastic order model.

Example 2.6 Consider $F(x) = \sqrt{x},\ x \in [0,1]$ and $G(x) = x,\ x \in [0,1]$. Then $F \leq_{st} G$. Hence, if we take G to play the role of F_0, then $\pi_0 = 0$ or, equivalently, $F \in SLB_0(G)$. On the other hand,

$$\pi_0' = \sup_{0<x<1} \frac{\sqrt{x} - x}{1 - x} = \frac{1}{2},$$

which means that it is necessary to trim at least 50% of the distribution F to make it stochastically larger than G, that is, if we want to see F as a contaminated version of a distribution stochastically larger than G, then the contamination must account, at least, for 50% of the distribution. With our notation, $F \in SUB_\pi(G)$ if and only if $\pi \geq \frac{1}{2}$.

If we exchange the roles of both distributions and take F to play the role of F_0, there is no trimming level $\pi < 1$, for which $(1-\pi)\sqrt{x} \leq x.\ x \in [0,1]$ holds. This means that $G \in SLB_\pi(F)$ is impossible for $\pi \in [0,1)$ and, consequently, we cannot see G as a contaminated version of distribution stochastically smaller than F.

Turning our focus now to the $\leq_{st}$ relation, we obviously have $F \leq_{st}^0 G$. Since $\sup_{x\in[0,1]} \sqrt{x} - x = 1/4$ we that $F \geq_{st}^\pi G$ if and only if $\pi \geq \frac{1}{4}$, that is, the minimum level of trimming in both distributions to reverse the original stochastic order is $\frac{1}{4}$. $\square$

Example 2.7 We take now $F(x) = \Phi(x-\mu)$ and $G(x) = \Phi(x)$, where Φ is the distribution function of the standard normal distribution, $N(0,1)$ and $\mu > 0$ (F is the distribution function of the $N(\mu,1)$ law). Obviously, $F \geq_{st} G$. Some calculus shows that $\sup_x \frac{\Phi(x)-\Phi(x-\mu)}{\Phi(x)} = 1$. Therefore, $F \in SLB_\pi(G)$ is impossible if $\pi < 1$ which means that the stochastic order between normal distributions with equal variances cannot be reversed by trimming one of them.

The picture is different when we move to the $\leq_{st}^\pi$ relation. It is easy to see that

$$\pi(F,G) = \sup_{x\in\mathbb{R}} (G(x) - F(x)) = \Phi\left(\tfrac{\mu}{2}\right) - \Phi\left(-\tfrac{\mu}{2}\right) = 2\Phi\left(\tfrac{\mu}{2}\right) - 1.$$

This means that for a shift of 0.1 units in location, we need trimming about 0.04 on both distributions to reverse the stochastic order. The required trimming is 0.0987 for a shift of 0.25 units, 0.1915 for a shift of 0.5 units and 0.3413 if the shift is of length one. $\square$

Remark 2.7.1 It is a well known fact that stochastic order is preserved by increasing transformations, namely, if $T : \mathbb{R} \to \mathbb{R}$ is an increasing function and $F \circ T^{-1}$ denotes the distribution function induced by F through T then $F_1 \leq_{st} F_2$ implies $F_1 \circ T^{-1} \leq_{st} F_2 \circ T^{-1}$. This carries over to the $\leq_{st}^\pi$ relation. In fact, let us assume that T is an increasing function. By Proposition 2.2 in Álvarez-Esteban et al. (2011)

$$\mathcal{R}_\pi(F \circ T^{-1}) = \{\tilde{F} \circ T^{-1},\ \tilde{F} \in \mathcal{R}_\pi(F)\}.$$

Preservation of the usual stochastic order implies that the transported probabilities $F^\pi \circ T^{-1}$ and $F_\pi \circ T^{-1}$ are the maximal and minimal, respectively, π-trimmed versions of

$F \circ T^{-1}$. In particular, this and Proposition 2.5 imply that if $F_1 \leq_{st}^{\pi} F_2$ then $F_1 \circ T^{-1} \leq_{st}^{\pi} F_2 \circ T^{-1}$. As a trivial consequence, for instance, if F and G are the distribution functions of lognormally distributed random variables $X = \exp(N + \mu)$ and $Y = \exp(N)$, where N is a standard normal random variable and $\mu > 0$, then $F \geq_{st} G$ and, by Example 2.7, $F \leq_{st}^{\pi} G$ if and only if $\pi \geq 2\Phi\left(\frac{\mu}{2}\right) - 1$. $\square$

We close this section with a comparison to alternative approaches to a *relaxed* version of the stochastic order. In Arcones et al (2002), the value

$$\theta(F, G) := P(X \leq Y) = \int (1 - G(x-))dF(x),$$

where X and Y are independent random variables with distribution functions F and G, is considered as an index of *precedence* of F to G. The relation $F \leq_{sp} G$ (F stochastically precedes to G) is defined by $\theta(F, G) \geq 1/2$. This is motivated by the fact that, if $F \leq_{st} G$ then $\theta(F, G) \geq 1/2$, with strict inequality unless $F = G$. On the other hand, this index can be greater than $1/2$, even if $F \not\leq_{st} G$. In other words, $F \leq_{st} G$ is a stronger relation than $F \leq_{sp} G$. It is argued then that the weaker nature of the relation $F \leq_{sp} G$ can be more versatile in some applications. As an illustrative example in Arcones et al (2002) the authors note that, if $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right), G(x) = \Phi\left(\frac{x-\nu}{\tau}\right)$ are the distribution functions of two normal laws, then $F \leq_{st} G$ if and only if $\mu \leq \nu$ and $\sigma = \tau$, while $F \leq_{sp} G$ as soon as $\mu \leq \nu$. However, we note that this precedence relation leads to consider that a distribution stochastically precedes that degenerated on its median, while, in fact, just half of the times it will produce values below its own median.

In contrast, and we think that more in line with the announced goal of giving a sound treatment to statements like 'We believe that population X is somehow smaller than population Y', we note that the relation $\leq_{st}^{\pi}$ allows to assess to what extent it can be expected that the values obtained from the first distribution will be smaller than those obtained from the second.

3 Inference in contaminated stochastic order models

In this section we will assume that $X_1, ..., X_n$ and $Y_1, ..., Y_m$ are independent i.i.d. random samples obtained from F and G, respectively. Our goal is to provide statistical methods for the assesment of contaminated stochastic order between F and G. More specifically, we are interested in testing the null model

$$H_0 : F \leq_{st}^{\pi} G, \tag{9}$$

for a fixed value of π against the alternative $H_a : F \not\leq_{st}^{\pi} G$. We are also interested in estimation of the minimal contamination level under which $F \leq_{st}^{\pi} G$ holds. The methods to be presented in this section could be easily adapted for inference about the π-level stochastic lower (or upper) bounds. For the sake of brevity we refrain from pursuing this issue in this paper.

From Proposition 2.5 it is clear that the testing problem and the estimation problem are very closely related, since we can rewrite (9) as the problem of testing

$$H_0 : \pi(F, G) \leq \pi \tag{10}$$

against the alternative $H_a : \pi(F, G) > \pi$. Therefore, if $\hat{L} = \hat{L}(X_1, ..., X_n, Y_1, ..., Y_m)$ were an (asymptotic) lower confidence bound for $\pi(F, G)$, rejection of H_0 when $\hat{L} > \pi$ would yield a test with (asymptotically) controlled type I error probability. We note also that often the real goal of the researcher will be *to conclude that stochastic order essentially holds*. But then, within the present setup, the testing problem under consideration should be

$$H_0 : \pi(F, G) \geq \pi \tag{11}$$

against $H_a : \pi(F, G) < \pi$. Rejection of (11) would provide statistical evidence that stochastic order, up to some (hopefully small) contamination, holds. In this case we could base our decission on an upper confidence bound for $\pi(F, G)$.

We will use the empirical version as the estimator of $\pi(F, G)$. More precisely, we write F_n and G_m, respectively, for the sample distribution functions of the X's and Y's samples and take

$$\pi(F_n, G_m) = \sup_{x \in \mathbb{R}} (G_m(x) - F_n(x))$$

as estimator of $\pi(F, G)$. We will make the following assumption in this section.

Assumption A1: F and G are continuous, and $n, m \to \infty$ in such a way that $\lambda_{n,m} := \frac{n}{n+m} \to \lambda \in (0, 1)$.

From the Glivenko-Cantelli theorem, we trivially obtain that $\pi(F_n, G_m)$ is a consistent estimator, namely,

$$\pi(F_n, G_m) \to \pi(F, G) \text{ almost surely, as } m, n \to \infty. \tag{12}$$

Under the homogeneity hypothesis $F = G$, it is well-known (see e.g. Durbin (1973)) that, if F and G are continuous, $\pi(F_n, G_m) = \sup_{x \in \mathbb{R}}(F_n(x) - G_m(x))$ is distribution free and, furthermore, that

$$\sqrt{\tfrac{mn}{m+n}} \pi(F_n, G_m) \to_w \bar{B} := \sup_{t \in [0,1]} B(t), \tag{13}$$

where $B(t)$ denotes a Brownian bridge on $[0, 1]$. This result allows easy computation of asymptotic critical values for testing the null model $F = G$. In fact, $P(\bar{B} > x) = \exp(-2x^2), x \geq 0$. On the other hand, for $F \neq G$, $\pi(F_n, G_m)$ is no longer distribution free, not even asymptotically as we can see in the next result. This suggests that we consider a bootstrap version of $\pi(F_n, G_m)$ as follows. Given $X_1, \ldots, X_n$ we take $X_1^*, \ldots, X_n^*$ to be i.i.d. with common ditribution F_n and write F_n^* for the empirical distribution on $X_1^*, \ldots, X_n^*$. Similarly we define G_m^*. With this setup we have the following.

Theorem 3.1 *Under assumption A1, if we denote* $\Gamma(F, G) := \{x \in \mathbb{R} : G(x) - F(x) = \pi(F, G)\}$ *and* $B_1(t)$ *and* $B_2(t)$ *are independent Brownian Bridges on* $[0, 1]$*, then*

$$\sqrt{\tfrac{mn}{m+n}}\,(\pi(F_n, G_m) - \pi(F, G)) \to_w \sup_{x\in\Gamma(F,G)} \left(\sqrt{\lambda}\ B_1(G(x)) - \sqrt{1-\lambda}\ B_2(F(x))\right). \tag{14}$$

Furthermore, if $\delta_{n,m} = K\sqrt{\frac{n+m}{nm}\log\log(\frac{nm}{n+m})}$*, with* $K > 2$*, and*

$$\Gamma_{n,m} = \{x :\ G_m(x) - F_m(x) \geq \pi(F_n, G_m) - \delta_{n,m}\}$$

then, conditionally given the X_i*'s and* Y_j*'s, with probability one,*

$$\sqrt{\tfrac{mn}{m+n}} \sup_{x\in\Gamma_{n,m}} ((G_m^*(x) - G_m(x)) - (F_n^*(x) - F_n(x)))$$
$$\to_w \sup_{x\in\Gamma(F,G)} \left(\sqrt{\lambda}\ B_1(G(x)) - \sqrt{1-\lambda}\ B_2(F(x))\right) \tag{15}$$

Remark 3.1.1 The convergence result in (14) is just a rewriting of Theorem 4 in Raghavachari (1973). In the Appendix we give a proof that yields (15) with little additional effort. We note that (14) includes (13) because for equal distributions $F = G$, we have $\pi(F, G) = 0$ and $\Gamma(F, G) = \mathbb{R}$. We observe further that with the alternative notation

$$T(F, G, \pi) := \{t \in [\pi, 1] :\ G(x) = t \text{ and } F(x) = t - \pi \text{ for some } x \in \mathbb{R}\},$$

the limit law in (14) can be rewritten

$$\bar{B}(F, G, \lambda) := \sup_{t\in T(F,G,\pi(F,G))} \left(\sqrt{\lambda}\ B_1(t) - \sqrt{1-\lambda}\ B_2(t - \pi(F, G))\right). \tag{16}$$

If $T(F, G, \pi(F, G))$ consists of a single point, say t_0 (note that $t_0 \in [\pi(F, G), 1]$ in that case), then $\bar{B}(F, G, \lambda)$ is centered normal with variance $\lambda t_0(1-t_0)+(1-\lambda)(t_0-\pi(F, G))(1-t_0+\pi(F, G))$. If $T(F, G, \pi(F, G))$ contains two or more points then $\bar{B}(F, G, \lambda)$ is not normal and has positive expectation. In fact, it is the possibility of having two or more points in $T(F, G, \pi(F, G))$ which has motivated the bootstrap proposal in Theorem 3.1 instead of simply considering the *direct* bootstrap version

$$\sqrt{\tfrac{mn}{m+n}}\,(\pi(F_n^*, G_m^*) - \pi(F_n, G_m))\,.$$

A look at the proof of Theorem 3.1 in the Appendix shows that the conditional asymptotic behavior of $\sqrt{\frac{mn}{m+n}}\,(\pi(F_n^*, G_m^*) - \pi(F_n, G_m))$ mimics that of $\sqrt{\frac{mn}{m+n}}(\pi(F_n, G_m) - \pi(F,\ G))$ (hence the bootstrap works) if $T(F, G, \pi(F, G))$ consists of only one point but can behave differently otherwise. See also Proposition 3.5 below.

It is convenient at this point to introduce the notation

$$\bar{B}(a, \lambda) := \sup_{t \in [a,1]} \left(\sqrt{\lambda}\ B_1(t) - \sqrt{1-\lambda}\ B_2(t-a) \right). \tag{17}$$

It is easy to see that $\bar{B}(a, \lambda)$ is a particular case of $\bar{B}(F, G, \lambda)$, coming, for instance, from the choice F (respectively G) equal to the distribution function of the uniform distribution on $[a, 1]$ (resp. uniform on $[0, 1]$). The next result provides simple but useful upper and lower bounds for the quantiles of $\bar{B}(F, G, \lambda)$.

Proposition 3.2 *If $\alpha \in (0, 1)$, $K_\alpha(F, G, \lambda)$ (resp. $K_\alpha(a, \lambda)$) is the α-quantile of $\bar{B}(F, G, \lambda)$ (resp. $\bar{B}(a, \lambda)$) and Φ denotes the standard normal distribution function then*

$$K_\alpha(F, G, \lambda) \leq K_\alpha(\pi(F, G), \lambda).$$

Furthermore, if $\alpha \in [\frac{1}{2}, 1)$

$$\bar{\sigma}(F, G, \pi(F, G))\Phi^{-1}(\alpha) \leq K_\alpha(F, G, \lambda)$$

where $\bar{\sigma}(F, G, \pi(F, G)) = \max_{t \in T(F,G,\pi(F,G))} \sigma_t$ and $\sigma_t^2 = \lambda t(1-t) + (1-\lambda)(t - \pi(F, G))(1 - t + \pi(F, G))$, while for $\alpha \in (0, \frac{1}{2})$

$$\underline{\sigma}(F, G, \pi(F, G))\Phi^{-1}(\alpha) \leq K_\alpha(F, G, \lambda)$$

with $\underline{\sigma}(F, G, \pi(F, G)) = \min_{t \in T(F,G,\pi(F,G))} \sigma_t$

From Proposition 3.2 we see that quantiles of $\bar{B}(F, G, \lambda)$ are bounded below by normal quantiles. Optimization of σ_t^2 over the interval $[\pi(F, G), 1]$ shows that

$$\underline{\sigma}_{\pi(F,G)} \leq \underline{\sigma}(F, G, \pi(F, G)) \leq \bar{\sigma}(F, G, \pi(F, G)) \leq \bar{\sigma}_{\pi(F,G)},$$

with $\underline{\sigma}_\pi^2 = \min(\lambda, 1-\lambda)\pi(1-\pi)$,

$$\bar{\sigma}_\pi^2 = \begin{cases} \frac{1}{4} - \pi^2\lambda(1-\lambda) & \text{if } \lambda\pi \leq \frac{1}{2} \text{ and } (1-\lambda)\pi \leq \frac{1}{2}, \\ \lambda\pi(1-\pi) & \text{if } \lambda\pi > \frac{1}{2}, \\ (1-\lambda)\pi(1-\pi) & \text{if } (1-\lambda)\pi > \frac{1}{2}. \end{cases} \tag{18}$$

This entails that for $\alpha \in [\frac{1}{2}, 1)$

$$\underline{\sigma}_{\pi(F,G)}\Phi^{-1}(\alpha) \leq K_\alpha(F, G, \lambda) \tag{19}$$

and for $\alpha \in (0, \frac{1}{2})$

$$\bar{\sigma}_{\pi(F,G)}\Phi^{-1}(\alpha) \leq K_\alpha(F, G, \lambda). \tag{20}$$

On the other hand, upper bounds for quantiles of $\bar{B}(F, G, \lambda)$ are given by quantiles of $\bar{B}(\pi(F, G), \lambda)$. On the other hand, upper bounds for quantiles of $\bar{B}(F, G, \lambda)$ are given by quantiles of $\bar{B}(\pi(F, G), \lambda)$. We provide next a useful expression for the computation of its quantiles. We also give a simple expression for the mean and the variance. Note that to avoid confusion we state the result for $\bar{B}(a, \lambda)$, with π denoting the usual constant in the following equations.

Proposition 3.3 *(a) If $a \in (0,1)$ and $u \in \mathbb{R}$*

$$P(\bar{B}(a,\lambda) > u\sqrt{1-a}) = 1 - \Phi(\tfrac{u}{\sqrt{\lambda a}})\Phi(\tfrac{u}{\sqrt{(1-\lambda)a}}) + e^{-\frac{2(1-a)u^2}{1-4\lambda(1-\lambda)a^2}}$$
$$\times \int_{-\infty}^{\frac{u}{\sqrt{\lambda a}}} \tfrac{1}{\sqrt{2\pi}} e^{-\frac{1-4\lambda(1-\lambda)a^2}{2}(x - \frac{2u\sqrt{\lambda a}(1-2(1-\lambda)a)}{1-4\lambda(1-\lambda)a^2})^2} \Phi\left(\tfrac{u(1-2(1-\lambda)a)}{\sqrt{(1-\lambda)a}} + 2\sqrt{\lambda(1-\lambda)}ax \right) dx.$$

(b) If $a \in [0,1)$

$$\begin{aligned} E(\bar{B}(a,\lambda)) &= \tfrac{1}{\sqrt{2\pi}}\left[\sqrt{a(1-a)} + \tfrac{\pi}{2} - \operatorname{atan}\left(\sqrt{\tfrac{a}{1-a}}\right)\right], \\ \operatorname{Var}(\bar{B}(a,\lambda)) &= \tfrac{1-a^2}{2} + \pi|2\lambda - 1|a - \tfrac{1}{2\pi}\left[\sqrt{a(1-a)} + \tfrac{\pi}{2} - \operatorname{atan}\left(\sqrt{\tfrac{a}{1-a}}\right)\right]^2. \end{aligned}$$

Part (a) of the last result easily yields that $\bar{B}(a,\lambda)$ has subgaussian tails, with, for instance, $P(\bar{B}(a,\lambda) > t) \leq 3e^{-\frac{2t^2}{1-4\lambda(1-\lambda)a^2}}$ for $t > 0$. It can also be used to compute approximately probabilities and quantiles of $\bar{B}(a,\lambda)$ through numerical integration. We return to this issue in Section 4. From (b) we see that $E(\bar{B}(a,\lambda))$ as a function of a (it does not depend on λ) decreases from $\sqrt{\frac{\pi}{8}}$, for $a = 0$ to 0 as $a \to 1$ and also that an unbalanced design ($\lambda \neq \frac{1}{2}$) results in an increase in variance, more important for large values of a.

3.1 Testing for essential stochastic order

We turn here to the testing problem (11), namely,

$$H_0 : \pi(F,G) \geq \pi_0 \quad \text{vs.} \quad H_a : \pi(F,G) < \pi_0 \tag{21}$$

for a fixed $\pi_0 \in (0,1)$. We reject H_0 for small values of $\pi(F_n, G_m)$. More precisely, if $\alpha < \frac{1}{2}$, our first proposal is rejection of H_0 in (21) if

$$\sqrt{\tfrac{nm}{n+m}}(\pi(F_n,G_m) - \pi_0) < \bar{\sigma}_{\pi_0}\Phi^{-1}(\alpha), \tag{22}$$

with $\bar{\sigma}_\pi$ as in (18). We show next that this provides a test of asymptotic level α, which detects alternatives with power exponentially close to one (see Remark 3.4.1 below). The result identifies a least favorable pair (other pairs are possible) within H_0. For a cleaner statement we will write $\pi_{m,n}$ for $\pi(F_n, G_m)$ and $\mathbb{P}_{F,G}$ for the probabilities under the assumption that the underlying distribution functions of the two samples are F and G, respectively.

Proposition 3.4 *With the above assumptions and notation, if $\alpha < \frac{1}{2}$ and $\pi_0 \leq \frac{1}{2}$,*

$$\begin{aligned} &\lim_{n\to\infty} \sup_{(F,G)\in H_0} \mathbb{P}_{F,G}\left[\sqrt{\tfrac{nm}{n+m}}(\pi_{n,m} - \pi_0) < \bar{\sigma}_{\pi_0}\Phi^{-1}(\alpha)\right] \\ &\quad = \lim_{n\to\infty} \mathbb{P}_{F_0,G_0}\left[\sqrt{\tfrac{nm}{n+m}}(\pi_{n,m} - \pi_0) < \bar{\sigma}_{\pi_0}\Phi^{-1}(\alpha)\right] = \alpha, \end{aligned} \tag{23}$$

where F_0 is the distribution function of the law $(\frac{1}{2}-\pi_0\lambda)U(0,\frac{1}{2}+\pi_0(1-\lambda))+(\frac{1}{2}+\pi_0\lambda)U(\frac{1}{2}+\pi_0(1-\lambda),1+\frac{\pi_0}{2}-\lambda\pi_0^2)$ and G_0 is the distribution function of the law $U(0,1)$. Furthermore, if $\pi(F,G)>\pi_0$ then

$$\mathbb{P}_{F,G}\left[\sqrt{\tfrac{nm}{n+m}}(\pi_{n,m}-\pi_0)<\bar{\sigma}_{\pi_0}\Phi^{-1}(\alpha)\right]\leq e^{-2\frac{nm}{n+m}(\pi_0-\pi(F,G))^2}, \tag{24}$$

while if $\pi(F,G)<\pi_0$ then for n,m such that $\frac{nm}{n+m}(\pi_0-\pi(F,G))^2\geq(\frac{1}{2}+\sqrt{\lambda_n(1-\lambda_n)})\log 2$ $-\bar{\sigma}_{\pi_0}\Phi^{-1}(\alpha)$, we have

$$\mathbb{P}_{F,G}\left[\sqrt{\tfrac{nm}{n+m}}(\pi_{n,m}-\pi_0)\geq\bar{\sigma}_{\pi_0}\Phi^{-1}(\alpha)\right]\leq 2e^{-\frac{2}{1+2\sqrt{\lambda_n(1-\lambda_n)}}(\bar{\sigma}_{\pi_0}\Phi^{-1}(\alpha)+\sqrt{\frac{nm}{n+m}}(\pi_0-\pi(F,G)))^2}. \tag{25}$$

Remark 3.4.1 Proposition 3.4 means that we can test $H_0:\pi(F,G)\geq\pi_0$ with a test of asymptotic level α and furthermore, that alternatives, $\pi(F,G)<\pi_0$, can be detected with power exponentially close to one. In fact, focusing for simplicity in the case $m=n$, we see from (25) that if $C<1$ then for large enough n any alternative $\pi(F,G)\leq\pi_1<\pi_0$ will be rejected with power at least $1-2e^{-\frac{C}{2}(\pi_0-\pi_1)^2n}$. If we combine this with (24) we see that our proposal yields a test of $H_0':\pi(F,G)\geq\pi_0'$ against $H_a':\pi(F,G)\leq\pi_1$ (with $\pi_0'>\pi_0>\pi_1$) which is uniformly exponentially consistent in the sense that both type I and type II error probabilities are uniformly exponentially small for large enough n. We refer to Barron (1989) for further discussion on uniformly exponentially consistent tests.

Remark 3.4.2 It is not really necessary to consider only the case $\pi_0\leq\frac{1}{2}$ in Proposition 3.4. In fact the same holds if $\pi_0\lambda\leq\frac{1}{2}$ and $\pi_0(1-\lambda)\leq\frac{1}{2}$, which is always the case if $\pi_0\leq\frac{1}{2}$. Otherwise, if $\pi_0>\frac{1}{2}$ we could have $\pi_0\lambda>\frac{1}{2}$ or $\pi_0(1-\lambda)>\frac{1}{2}$. In the first case Proposition 3.4 holds if we take F_0 to be the distribution function of the law $U(\pi_0,1)$ and in the second we have to take the law $(1-\pi_0)U(0,1)+\pi_0U(1,1+\pi_0(1-\pi_0))$. Details can be checked in a straighforward way. From an applied point of view the interest of Proposition 3.4 is to assess that stochastic order holds up to some small contamination, say $\pi_0=0.1$, $\pi_0=0.05$ or $\pi_0=0.01$. For this reasons and to get a simpler statement we have chosen to focus on the case $\pi_0\leq\frac{1}{2}$.

While Proposition 3.4 guarantees fast convergence to 0 of type II error probabilities and of type I error probabilities as we move away from the boundary with the test in (22), the test is somewhat conservative for finite samples as we will see in Section 4 and some alternative procedures can be of interest. One possibility is to reject $H_0:\pi(F,G)\geq\pi_0$ if

$$\sqrt{\tfrac{nm}{n+m}}(\pi_{n,m}-\pi_0)<\hat{\sigma}_{n,m}\Phi^{-1}(\alpha), \tag{26}$$

where $\hat{\sigma}_{n,m}^2=\min_{t\in T(F_n,G_m,\pi(F_n,G_m))}\sigma_t^2$ and σ_t^2 is as in Proposition 3.2 replacing F with F_n and G with G_m. A little thought shows that this increases (slightly) the probability of rejection at the boundary (if $\alpha < \frac{1}{2}$ we are increasing the cut value), while providing still a test of asymptotic level α. In Section 4 we show the improvement that (26) provides over (22) for finite samples.

A second, more important source of improvement comes from the consideration of bias corrected estimators instead of $\pi(F_n, G_m)$. In fact

$$E(\sup_x (G_m(x) - F_n(x)) \geq \sup_x E(G_m(x) - F_n(x)) = \sup_x (G(x) - F(x)).$$

This implies $\text{bias}(\pi(F_n, G_m)) = E(\pi(F_n, G_m)) - \pi(F, G) \geq 0$. Furthermore, it is easy to see from the proofs in the Appendix that

$$\sqrt{\tfrac{mn}{m+n}}\text{bias}(\pi(F_n, G_m)) \to E\Big(\sup_{t \in T(F,G,\pi(F,G))} \sqrt{\lambda} B_1(t) - \sqrt{1-\lambda} B_2(t - \pi(F, G))\Big).$$

Combining this last display with (b) in Proposition 3.3 and subsequent comments, we see that, asymptotically, $\text{bias}(\pi(F_n, G_m))$ is smaller than $\sqrt{\frac{\pi}{8}}\sqrt{\frac{m+n}{mn}} \simeq 0.63\sqrt{\frac{m+n}{mn}}$. We also see that $\sqrt{\frac{m+n}{mn}}\text{bias}(\pi(F_n, G_m)) \to 0$ when $\pi(F_n, G_m)$ is asymptotically normal (that is, when $T(F, G, \pi(F, G))$ consists of a single point).

We consider the bootstrap bias estimator

$$\widehat{\text{bias}}_{\text{BOOT}}(\pi(F_n, G_m)) := E^*(\pi(F_n^*, G_m^*)) - \pi(F_n, G_m),$$

where F_n^*, G_m^* are as in Theorem 3.1 and E^* denotes conditional expectation given the X_i's and Y_j's. We define the bias corrected estimator

$$\hat{\pi}_{n,m,\text{BOOT}} := \pi(F_n, G_m) - \widehat{\text{bias}}_{\text{BOOT}}(\pi(F_n, G_m)). \tag{27}$$

The next result is the key for the performance of $\hat{\pi}_{n,m,\text{BOOT}}$.

Proposition 3.5 *Under the assumptions of Theorem 3.1 we have*

$$\sqrt{\tfrac{mn}{m+n}}\widehat{\text{bias}}_{\text{BOOT}}(\pi(F_n, G_m)) \to 0.$$

in probability as $n, m \to \infty$.

A proof is given in the Appendix. Proposition 3.5 shows that $\sqrt{\frac{mn}{m+n}}(\pi(F_n^*, G_m^*) - \pi(F_n, G_m))$ does not mimic the asymptotic behavior of $\sqrt{\frac{mn}{m+n}}(\pi(F_n, G_m) - \pi(F, G))$ unless $\pi(F_n, G_m)$ is asymptotically normal. But, more importantly, it shows that the bootstrap bias correction does not affect the first order behavior of $\pi(F_n, G_m)$. In other words, that rejection of $H_0 : \pi(F, G) \geq \pi_0$ if

$$\sqrt{\tfrac{nm}{n+m}}(\hat{\pi}_{n,m,\text{BOOT}} - \pi_0) < \hat{\sigma}_{n,m}\Phi^{-1}(\alpha), \tag{28}$$

with $\hat{\sigma}_{n,m}$ as above, is a test of asymptotic level α with fastly decreasing type I and type II error probabilities away from the null hypothesis boundary. We show later, in a simulation study in Section 4, that the bootstrap correction (which of course has to be approximated, in turn, through bootstrap simulation) yields a significant improvement with respect to (22) or (26) in terms of power and approximation of the nominal level for finite samples.

3.2 Testing against essential stochastic order

In some instances the researcher can be interested in gathering statistical evidence against stochastic order, or to stochastic order up to some small contamination. The relevant testing problem is then (10), namely,

$$H_0 : \pi(F,G) \leq \pi_0 \tag{29}$$

against the alternative $H_a : \pi(F,G) > \pi_0$. Now, we would reject the null hypothesis for large values of $\pi(F_n, G_m)$. Motivated by Proposition 3.2 we consider the test that rejects H_0 in (29) if

$$\sqrt{\tfrac{nm}{n+m}}(\pi_{n,m} - \pi_0) > K_{1-\alpha}(\pi_0, \lambda), \tag{30}$$

where $K_{1-\alpha}(\pi_0, \lambda)$ is the $1-\alpha$ quantile of $\bar{B}(\pi_0, \lambda)$ defined in (17). Next, we give the main facts about the test (30). As in the statement of Proposition 3.4 we write $\pi_{m,n}$ for $\pi(F_n, G_m)$ and $\mathbb{P}_{F,G}$ for the probabilities under the assumption that the underlying distribution functions of the two samples are F and G, respectively.

Proposition 3.6 *With the above assumptions and notation,*

$$\begin{aligned}\lim_{n\to\infty} \sup_{(F,G)\in H_0} \mathbb{P}_{F,G}\left[\sqrt{\tfrac{nm}{n+m}}(\pi_{n,m} - \pi_0) > K_{1-\alpha}(\pi_0,\lambda)\right] \\ = \lim_{n\to\infty} \mathbb{P}_{F_0,G_0}\left[\sqrt{\tfrac{nm}{n+m}}(\pi_{n,m} - \pi_0) > K_{1-\alpha}(\pi_0,\lambda)\right] = \alpha,\end{aligned} \tag{31}$$

where F_0 is the distribution function of the law $U(\pi_0, 1+\pi_0)$ and G_0 is the distribution function of the law $U(0,1)$. Furthermore, if $\pi(F,G) < \pi_0$ and $K_{1-\alpha}(\pi_0,\lambda) \geq 0$ then

$$\mathbb{P}_{F,G}\left[\sqrt{\tfrac{nm}{n+m}}(\pi_{n,m} - \pi_0) > K_{1-\alpha}(\pi_0,\lambda)\right] \leq 2e^{-\frac{2}{1+2\sqrt{\lambda_n(1-\lambda_n)}}\frac{nm}{n+m}(\pi_0 - \pi(F,G))^2}, \tag{32}$$

while if $\pi(F,G) > \pi_0$ then

$$\mathbb{P}_{F,G}\left[\sqrt{\tfrac{nm}{n+m}}(\pi_{n,m} - \pi_0) \leq K_{1-\alpha}(\pi_0,\lambda)\right] \leq e^{-2(\sqrt{\frac{nm}{n+m}}(\pi_0 - \pi(F,G)) - K_{1-\alpha}(\pi_0,\lambda))^2}. \tag{33}$$

A proof of Proposition 3.6 is given in the Appendix. Similar comments as in Remark 3.4.1 can be made now. The test in (30) is asymptotically of level α for $H_0 : \pi(F,G) \leq \pi_0$ vs. $H_a : \pi(F,G) > \pi_0$ and uniformly exponentially consistent test for $H_0' : \pi(F,G) \leq \pi_0'$ vs. $H_a' : \pi(F,G) > \pi_1$ if $\pi_0' < \pi_0 < \pi_1$. Later, in Section 4 we will see that this test shows a good performance for finite sample sizes (even for small sizes).

3.3 Confidence bounds.

Rather than testing for or against the contaminated stochastic order model one could prefer to report results in terms of confidence intervals for the true contamination level, $\pi(F,G)$. Here we discuss briefly upper and lower confidence bounds for $\pi(F,G)$. Proper two-sided confidence intervals can we constructed from our confidence bounds in a straighforward way. We omit details.

Recalling Theorem 3.1 and Proposition 3.2 we see that

$$\pi(F_n,G_m)-\sqrt{\tfrac{n+m}{nm}}K_\alpha(F,G,\lambda),$$

is an ideal upper confidence bound, asymptotically of level $1-\alpha$ for $\pi(F,G)$, that cannot be used directly, since the quantiles $K_\alpha(F,G,\lambda)$ are unknown. It follows from Theorem 3.1 that $K_\alpha(F,G,\lambda)$ can be consistently estimated by the conditional α-quantile of $\sqrt{\frac{mn}{m+n}}\sup_{x\in\Gamma_n(F_n,G_m)}((G_m^*(x)-G_m(x))-(F_n^*(x)-F_n(x)))$, that we denote by $\hat{K}_\alpha^{(\mathrm{Boot})}$ which can be approximated by simulation. As a result, we have that

$$\pi(F_n,G_m)-\sqrt{\tfrac{n+m}{nm}}\hat{K}_\alpha^{(\mathrm{Boot})} \tag{34}$$

is an upper confidence bound for $\pi(F,G)$ with asymptotic confidence level $1-\alpha$. Unfortunately, our simulations show that the finite sample performance of this upper confidence bound or test can be too liberal even for large sample sizes. Hence it can be better to consider different confidence bounds.

Assuming $\alpha<\frac{1}{2}$, it follows from Propositions 3.2 and 3.5 that

$$\hat{\pi}_{n,m,\mathrm{BOOT}}-\sqrt{\tfrac{n+m}{nm}}\hat{\sigma}_{m,n}\Phi^{-1}(\alpha) \tag{35}$$

with $\hat{\pi}_{n,m,\mathrm{BOOT}}$ and $\hat{\sigma}_{m,n}$ as in (28) is an upper bound with asymptotic confidence level at least $1-\alpha$. Our simulations in Section 4 show a good performance of (35) for finite samples.

Turning to the issue of lower confidence bounds, Theorem 3.1 and Proposition 3.2 imply that

$$\pi(F_n,G_m)-\sqrt{\tfrac{n+m}{nm}}K_{1-\alpha}(\pi(F_n,G_m),\lambda_{m,n}) \tag{36}$$

is a lower confidence bound for $\pi(F,G)$ with asymptotic confidence level $1-\alpha$. As for the test in (30), quantiles $K_{1-\alpha}(\pi(F_n,G_m),\lambda_{m,n})$ can be numerically approximated from part (a) of Proposition 3.3.

4 Simulations and Case Study

We explore here the finite sample performance of the tests and confidence bounds introduced in Section 3. We start with the tests for essential stochastic order (22), (26) and

(28). We consider several values of π_0 and have simulated from different pairs (F, G). Proposition 3.4 tells us that (at least asymptotically) for a fixed value of $\pi = \pi(F, G)$, type I error probability is largest for $F_{\pi,b}$ corresponding to $(\frac{1}{2} - \pi\lambda)U(0, \frac{1}{2} + \pi(1-\lambda)) + (\frac{1}{2} + \pi\lambda)U(\frac{1}{2} + \pi(1-\lambda), 1 + \frac{\pi}{2} - \lambda\pi^2)$ and G coming from the uniform law on $(0, 1)$, while from the point of view of power the worst performance (recall Theorem 3.1 and Proposition 3.2) should come from the pair $(F_{\pi,a}, G)$ with $F_{\pi,a}$ the d.f. of the uniform law on $(\pi, 1+\pi)$ and G as before. Consequently, we have simulated samples from these choices $F_{\pi,a}$, $F_{\pi,b}$ and G for several values of π. We have also considered the case $F_0 = G$. Although we have some indication that the balance of sample sizes has an impact on the performance of the procedure (see the comments after Proposition 3.3) we have, for the sake of brevity, focused on the case $m = n$ and have considered different sample sizes. In the next tables we show the simulated rejection frequencies for the tests (22), (26) and (28). In all cases we have computed this simulated rejection frequency from 1000 replicates of the procedure. In the case of test (28) the bootstrap bias correction has been approximated by the average from 1000 bootstrap replicates. In all cases the nominal level of the test was $\alpha = 0.05$ and G is the d.f. of the uniform law on $(0, 1)$.

Table 1: Observed rejection frequencies. $H_0 : \pi(F, G) \geq \pi_0$ vs. $H_a : \pi(F, G) < \pi_0$
$G = U(0, 1)$, $m = n$; reject if $\sqrt{\frac{nm}{n+m}}(\pi_{n,m} - \pi_0) < \bar{\sigma}_{\pi_0}\Phi^{-1}(0.05)$

$\boldsymbol{\pi_0}$	$\boldsymbol{n}$	$\boldsymbol{F_{0.2,a}}$	$\boldsymbol{F_{0.2,b}}$	$\boldsymbol{F_{0.1,a}}$	$\boldsymbol{F_{0.1,b}}$	$\boldsymbol{F_{0.05,a}}$	$\boldsymbol{F_{0.05,b}}$	$\boldsymbol{F_{0.01,a}}$	$\boldsymbol{F_{0.01,b}}$	$\boldsymbol{F_0}$
0.01	50	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	1000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	5000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	10000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.05	50	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	1000	0.000	0.000	0.000	0.000	0.000	0.000	0.016	0.071	0.183
	5000	0.000	0.000	0.000	0.000	0.000	0.008	0.924	0.939	0.995
	10000	0.000	0.000	0.000	0.000	0.000	0.021	0.999	1.000	1.000
0.1	50	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	500	0.000	0.000	0.000	0.001	0.010	0.152	0.532	0.589	0.698
	1000	0.000	0.000	0.000	0.011	0.155	0.491	0.946	0.952	0.979
	5000	0.000	0.000	0.000	0.025	0.996	1.000	1.000	1.000	1.000
	10000	0.000	0.000	0.000	0.036	1.000	1.000	1.000	1.000	1.000
0.2	50	0.000	0.000	0.000	0.004	0.006	0.024	0.045	0.048	0.075
	100	0.000	0.002	0.010	0.087	0.151	0.296	0.480	0.499	0.559
	500	0.000	0.021	0.706	0.880	0.997	0.996	1.000	1.000	1.000
	1000	0.000	0.022	0.985	0.994	1.000	1.000	1.000	1.000	1.000
	5000	0.000	0.047	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	10000	0.000	0.041	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 2: Observed rejection frequencies. $H_0 : \pi(F,G) \geq \pi_0$ vs. $H_a : \pi(F,G) < \pi_0$
$G = U(0,1)$, $m = n$; reject if $\sqrt{\frac{nm}{n+m}}(\pi_{n,m} - \pi_0) < \hat{\sigma}_{m,n}\Phi^{-1}(0.05)$

π_0	n	$F_{0.2,a}$	$F_{0.2,b}$	$F_{0.1,a}$	$F_{0.1,b}$	$F_{0.05,a}$	$F_{0.05,b}$	$F_{0.01,a}$	$F_{0.01,b}$	F_0
0.01	50	0.000	0.000	0.000	0.000	0.000	0.003	0.007	0.015	0.017
	100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.016
	500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.007
	1000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.016
	5000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.051
	10000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.023	0.117
0.05	50	0.000	0.000	0.000	0.001	0.000	0.003	0.005	0.011	0.017
	100	0.000	0.000	0.000	0.001	0.000	0.007	0.020	0.036	0.039
	500	0.000	0.000	0.000	0.000	0.000	0.015	0.100	0.105	0.147
	1000	0.000	0.000	0.000	0.000	0.000	0.022	0.206	0.231	0.360
	5000	0.000	0.000	0.000	0.000	0.000	0.023	0.948	0.973	0.998
	10000	0.000	0.000	0.000	0.000	0.000	0.025	1.000	1.000	1.000
0.1	50	0.000	0.000	0.000	0.011	0.010	0.042	0.063	0.064	0.080
	100	0.000	0.000	0.000	0.011	0.019	0.033	0.123	0.124	0.137
	500	0.000	0.000	0.000	0.017	0.151	0.216	0.606	0.693	0.761
	1000	0.000	0.000	0.000	0.025	0.337	0.523	0.961	0.957	0.979
	5000	0.000	0.000	0.000	0.028	0.993	0.999	1.000	1.000	1.000
	10000	0.000	0.000	0.000	0.033	1.000	1.000	1.000	1.000	1.000
0.2	50	0.000	0.015	0.048	0.101	0.151	0.171	0.273	0.295	0.288
	100	0.000	0.013	0.118	0.175	0.369	0.406	0.580	0.614	0.643
	500	0.000	0.020	0.770	0.880	0.993	0.993	1.000	0.999	1.000
	1000	0.000	0.033	0.988	0.995	1.000	1.000	1.000	1.000	1.000
	5000	0.000	0.038	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	10000	0.000	0.032	1.000	1.000	1.000	1.000	1.000	1.000	1.000

We see in Table 1 how alternatives are detected with power rapidly increasing to 1, as predicted by (25). For instance, in this balanced setup ($m = n$), if we fix $\pi_0 = 0.05$ (we are trying to establish that F is stochastically smaller than G up to 5% contamination) then, to guarantee that alternatives with $\pi(F,G) = 0.01$ are detected with power at least 90% the bound (25) requires a sample size $n = m = 8143$. In the simulation study we observe that the power is above 90% for $n = m = 5000$. We also see the very small type I error probability guaranteed by (24). In fact, we see that the test in (22) is somewhat conservative for finite samples with slow convergence to the nominal level This is more clearly seen for small values of π_0. In Table 2 we see that the correction (26) improves slightly the convergence to the nominal level, resulting in some increase in power while keeping the low type I error probabilities. Table 3 shows the remarkable effect of the bootstrap bias correction (28). We see that sample sizes about $n = m = 1000$ suffice to give a rather close agreement to the nominal level, even for small values of π_0. And we also see that the bias correction results in a significant increase in power. As an example, if we are trying to reject that there is more than 10% contamination with respect to the stochastic order model and we were, in fact, sampling from distributions with 5%

Table 3: Observed rejection frequencies. $H_0 : \pi(F,G) \geq \pi_0$ vs. $H_a : \pi(F,G) < \pi_0$
$G = U(0,1)$, $m = n$; reject if $\sqrt{\frac{nm}{n+m}}(\hat{\pi}_{n,m,\mathrm{BOOT}} - \pi_0) < \hat{\sigma}_{n,m}\Phi^{-1}(0.05)$

$\boldsymbol{\pi_0}$	$\boldsymbol{n}$	$\boldsymbol{F_{0.2,a}}$	$\boldsymbol{F_{0.2,b}}$	$\boldsymbol{F_{0.1,a}}$	$\boldsymbol{F_{0.1,b}}$	$\boldsymbol{F_{0.05,a}}$	$\boldsymbol{F_{0.05,b}}$	$\boldsymbol{F_{0.01,a}}$	$\boldsymbol{F_{0.01,b}}$	$\boldsymbol{F_0}$
0.01	50	0.000	0.000	0.000	0.003	0.001	0.012	0.019	0.034	0.038
	100	0.000	0.000	0.000	0.001	0.000	0.005	0.015	0.027	0.028
	500	0.000	0.000	0.000	0.000	0.000	0.001	0.016	0.028	0.049
	1000	0.000	0.000	0.000	0.000	0.000	0.000	0.013	0.030	0.063
	5000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.038	0.136
	10000	0.000	0.000	0.000	0.000	0.000	0.000	0.013	0.051	0.277
0.05	50	0.000	0.000	0.000	0.009	0.015	0.042	0.062	0.067	0.093
	100	0.000	0.000	0.000	0.006	0.008	0.038	0.065	0.098	0.106
	500	0.000	0.000	0.000	0.003	0.007	0.058	0.208	0.220	0.332
	1000	0.000	0.000	0.000	0.001	0.009	0.039	0.415	0.426	0.566
	5000	0.000	0.000	0.000	0.000	0.003	0.057	0.978	0.987	1.000
	10000	0.000	0.000	0.000	0.000	0.006	0.050	1.000	1.000	1.000
0.1	50	0.000	0.005	0.003	0.030	0.054	0.089	0.137	0.138	0.134
	100	0.000	0.001	0.007	0.052	0.076	0.121	0.246	0.250	0.266
	500	0.000	0.000	0.007	0.040	0.337	0.387	0.801	0.830	0.876
	1000	0.000	0.000	0.005	0.056	0.589	0.661	0.976	0.985	0.997
	5000	0.000	0.000	0.003	0.057	0.999	1.000	1.000	1.000	1.000
	10000	0.000	0.000	0.008	0.058	1.000	1.000	1.000	1.000	1.000
0.2	50	0.007	0.051	0.130	0.172	0.321	0.350	0.483	0.469	0.508
	100	0.003	0.068	0.280	0.339	0.541	0.600	0.758	0.761	0.798
	500	0.004	0.051	0.888	0.928	1.000	0.997	1.000	1.000	1.000
	1000	0.002	0.050	0.999	0.999	1.000	1.000	1.000	1.000	1.000
	5000	0.002	0.054	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	10000	0.004	0.045	1.000	1.000	1.000	1.000	1.000	1.000	1.000

contamination or less, then samples of size 1000 would give a probability of rejection of 60% or more and from samples of size 5000 we would reject with probability close to 1. Even for the hard problem of concluding that F and G satisfy the stochastic order up to 1% contamination we see nonnegligible power for $n = 5000$ or 10000, a sample size not unusual in some econometric studies (for instance, the Canadian Family Expenditure Survey, considered in Barrett and Donald (2003) involves more than 9000 units).

In the testing problem (29), namely, the problem of looking for statistical evidence against stochastic order up to some small contamination, we have considered the test (30), that is, $H_0 : \pi(F,G) \leq \pi_0$ is rejected if $\sqrt{\frac{n}{2}}(\pi(F_n, G_n) - \pi_0) > K_{1-\alpha}(\pi_0, \frac{1}{2})$. $K_{1-\alpha}$ has been aproximated using the expresion in Proposition 3.3 (a) plus numerical integration and inversion. As before, we have focused on the case $m = n$ and $\alpha = 0.05$. In this case, for a fixed value of $\pi = \pi(F,G)$, the worst case from the point of view of type I error corresponds to $\tilde{F}_{\pi,a} = U(\pi, 1+\pi)$ vs $G = U(0,1)$ while the worst case for power is $\tilde{F}_{\pi,b} = \frac{1-\pi}{2}U(0, \frac{1+\pi}{2}) + \frac{1+\pi}{2}U(\frac{1+\pi}{2}, 1+\frac{\pi(1-\pi)}{2})$ vs. $G = U(0,1)$ and these are the distributions that we have considered. Again, we have also considered $\tilde{F}_0 = U(0,1)$. The results are reported in Table 4. We observe a very good agreement between nominal and simulated levels, even for small values of n. We also see rapidly decaying error probabilites as predicted by (32) and (33).

Table 4: Observed rejection frequencies. $H_0 : \pi(F, G) \leq \pi_0$ vs. $H_a : \pi(F, G) > \pi_0$
$G = U(0, 1)$, $m = n$; reject if $\sqrt{\frac{nm}{n+m}}(\pi_{n,m} - \pi_0) > K_{0.95}(\pi_0, \lambda_{n,m})$

π_0	n	$\tilde{F}_0$	$\tilde{F}_{0.01,a}$	$\tilde{F}_{0.01,b}$	$\tilde{F}_{0.05,a}$	$\tilde{F}_{0.05,b}$	$\tilde{F}_{0.1,a}$	$\tilde{F}_{0.1,b}$	$\tilde{F}_{0.2,a}$	$\tilde{F}_{0.2,b}$
0.01	50	0.045	0.039	0.052	0.060	0.111	0.159	0.302	0.485	0.824
	100	0.022	0.031	0.040	0.066	0.107	0.199	0.450	0.848	0.986
	500	0.021	0.026	0.045	0.210	0.477	0.951	1.000	1.000	1.000
	1000	0.011	0.028	0.047	0.441	0.774	1.000	1.000	1.000	1.000
	5000	0.002	0.017	0.049	1.000	1.000	1.000	1.000	1.000	1.000
	10000	0.001	0.005	0.047	1.000	1.000	1.000	1.000	1.000	1.000
0.05	50	0.015	0.010	0.016	0.023	0.052	0.056	0.142	0.253	0.600
	100	0.004	0.007	0.009	0.014	0.031	0.069	0.186	0.518	0.885
	500	0.000	0.001	0.002	0.009	0.047	0.211	0.664	1.000	1.000
	1000	0.000	0.000	0.000	0.001	0.060	0.606	0.954	1.000	1.000
	5000	0.000	0.000	0.000	0.001	0.040	1.000	1.000	1.000	1.000
	10000	0.000	0.000	0.000	0.000	0.056	1.000	1.000	1.000	1.000
0.1	50	0.001	0.002	0.005	0.004	0.007	0.009	0.027	0.079	0.274
	100	0.000	0.003	0.000	0.001	0.002	0.010	0.031	0.163	0.520
	500	0.000	0.000	0.000	0.000	0.001	0.002	0.056	0.958	0.999
	1000	0.000	0.000	0.000	0.000	0.000	0.001	0.035	1.000	1.000
	5000	0.000	0.000	0.000	0.000	0.000	0.000	0.056	1.000	1.000
	10000	0.000	0.000	0.000	0.000	0.000	0.000	0.057	1.000	1.000
0.2	50	0.000	0.000	0.000	0.000	0.000	0.001	0.004	0.004	0.029
	100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.051
	500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.044
	1000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.052
	5000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.053
	10000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.040

Now we will discuss on the evolution associated to age of the heights of boys and girls. It is a commonly assumed fact that generally boys are taller than girls but also that girls are more precocious in body development than boys. We analyze these facts through our methodology on the basis of a data set obtained from NHANES (National Health and Nutrition Examination Survey). We will consider just the more interesting case of obtaining upper bounds for the levels of stochastic order in both senses. Data are obtained from the surveys of the years 1999, 2001, 2003, 2005, 2007 and 2009. We have considered individuals in growth age, from 2 to 19 years. The sample sizes for every cohort by sex and age are showed in Table 5, the smallest global sample size corresponding to the age of 10 years (556 boys and 536 girls).

Through this analysis F_n^a and G_m^a will respectively denote the empirical distribution functions (EDF's) of girls and boys at age$=a$.

The graphics in Figure 1 show some illustrative situations of the joint behavior of

Table 5: Sample sizes by age (boys, top row; girls, bottom row).

2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
796	632	633	563	557	582	579	543	556	556	735	728	704	716	793	783	716	725
776	563	620	567	542	564	572	579	536	587	733	757	764	665	702	703	716	647

the empirical distribution functions (EDF). As a global summary, Figure 2 shows the evolution of $\pi(F_n^a, G_m^a)$ (red) and $\pi(G_m^a, F_n^a)$ (green) for $a = 2, ..., 19$. The application of (35), with $\alpha = 0.05$ and 1000 bootstrap samples for every year, allow to produce the sequence of upper bounds for $\pi(F^a, G^a)$ and $\pi(G^a, F^a)$ (F^a and G^a are the parent distribution functions) with asymptotic confidence level of 95% that are represented in Figure 3.

From the obtained bounds (see table 6 below), we would obtain statistical evidence enough to confirm that boys are essentially taller than girls until 8 years and from 14 onwards. However, at 10 or 11 years old, the girls are essentially taller than boys (for instance, with confidence 95% girls are taller than boys at age 11 except for a contamination lower than 2%).

Table 6: 95%-Upper bounds by age for $\pi(F^a, G^a)$ (top row) and $\pi(G^a, F^a)$ (bottom).

2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0.01	0.02	0.01	0.02	0.02	0.06	0.04	0.08	0.30	0.30	0.29	0.08	0.01	0.01	0	0	0	0
0.33	0.16	0.27	0.32	0.28	0.21	0.24	0.13	0.04	0.02	0.12	0.56	0.91	1.00	1	1	1	1

5 Discussion

In the stochastic dominance setting, the approximation to the model given by the mixture approach can be easily characterized and strongly suggests the estimation of the mixture index as well as the testing statistics. In fact this mixture index gives a nice interpretation to the meaning of the one-sided Kolmogorov-Smirnov statistic.This is a consequence of the fact that although the contaminating distributions can distort the model in very different ways, there exist a most unfavorable way for that, thus the less favorable hypothesis can be stated in terms of just two distributions involving the parent distributions P_1 and P_2 and the maximum level of contamination allowed, say π_0.

The introduced methodology is based on the consideration of the natural estimator of the contamination level given by the empirical level $\pi(F_n, G_m)$. In the paper we have showed the feasibility of this methodology, giving non-asymptotic computable bounds and justifying the use of bootstrap approximations and bias corrections. Our approach allows to address the statistical assessment of essential stochastic dominance giving a sense to the meaning of "essential" through the level $\pi(F, G)$ (problem c) in the introduction). Since this is the, so far, less explored problem, we have considered it in more detail in

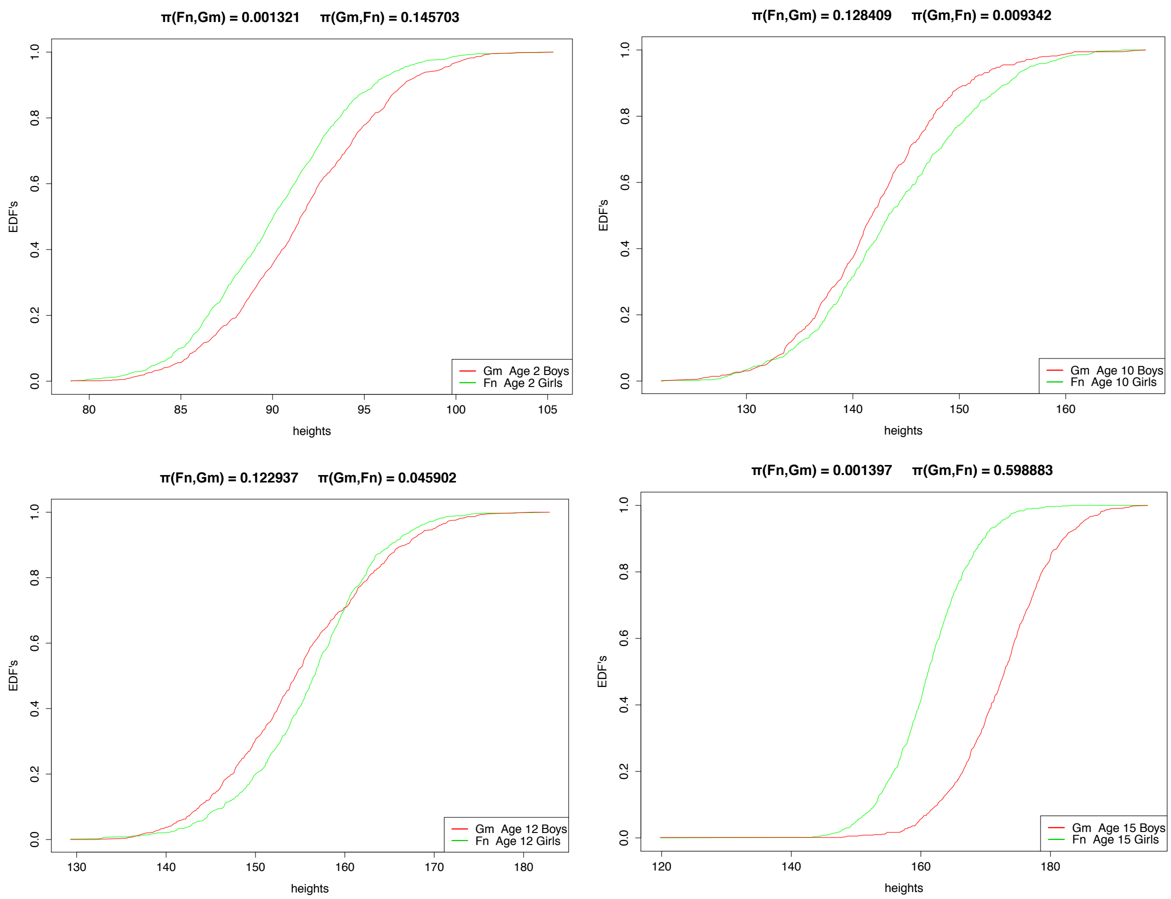

Figure 1: Graphics (corresponding to ages 2, 10, 12 and 15 years) showing the different evolution of the EDF's for boys and girls. The main legends give the values $\pi(F_n^a, G_m^a)$ and $\pi(G_m^a, F_n^a)$ for these years.

the applications, but our results also cover and extend the framework of testing against stochastic dominance.

Finally we should highlight that the level $\pi(F_n, G_m)$ can be also used as a descriptive index of the evolution in time of an stochastic dominance. In a forthcoming paper we will explore the possibilities in the multivariate or in the stochastic processes framework.

Appendix

In this Appendix we provide proofs for the results in Section 3. Most of them are related to the behavior of $\pi(F_n, G_m)$. We keep the notation of Section 3, including that for the

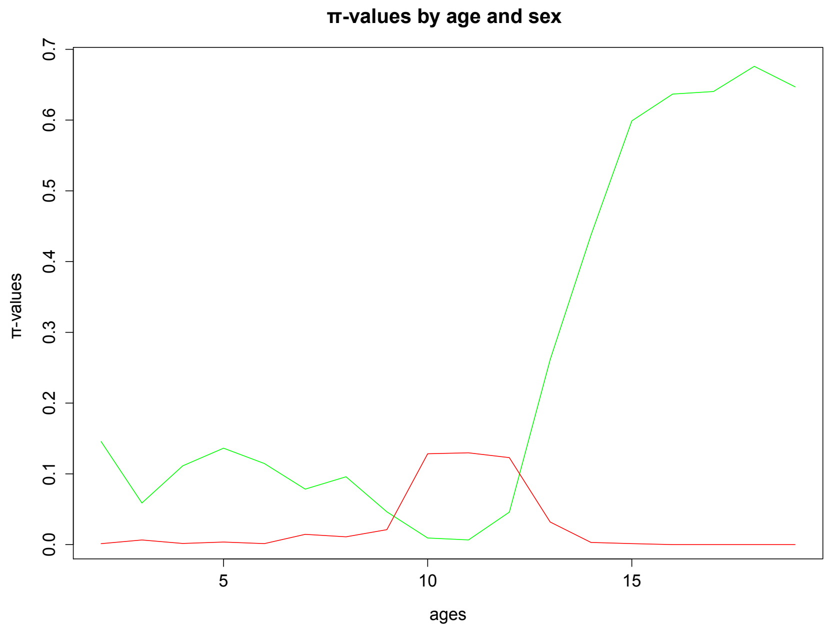

Figure 2: Evolution of $\pi(F_n^a, G_m^a)$ (red) and $\pi(G_m^a, F_n^a)$ (green) for $a = 2, ..., 19$.

sets

$$\Gamma(F,G) = \{x \in \bar{\mathbb{R}} : G(x) - F(x) = \pi(F,G)\}$$

and

$$T(F,G,\pi) = \{t \in [0,1] : G(x) = t, F(x) = t - \pi \text{ for some } x \in \bar{\mathbb{R}}\}$$

(we write $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ in the definition of the sets, with $G(+\infty) - F(+\infty) = G(-\infty) - F(-\infty) = 0$, to cover the case $\pi(F,G) = 0$ with $G(x) < F(x)$ for all $x \in \mathbb{R}$).

Throughout this Appendix we will assume (without loss of generality) that $X_i = F^{-1}(U_i)$, $i = 1, \ldots, n$ and $Y_j = G^{-1}(V_j)$, $j = 1, \ldots, m$ where $U_1, \ldots, U_n, V_1, \ldots, V_m$ are i.i.d. r.v.'s. We will write $\alpha_{m,1}(t)$, $0 \le t \le 1$ for the empirical process based on the V_j's and $\alpha_{n,2}$ for the empirical process on the U_i's. We note that, in particular, $G_m(x) = G(x) + \frac{1}{\sqrt{m}}\alpha_{m,1}(G(x))$ and $F_n(x) = F(x) + \frac{1}{\sqrt{n}}\alpha_{n,2}(F(x))$. We will use this fact throughout this Appendix without further mention. We introduce the processes

$$\alpha_{n,m}(s,t) = \sqrt{\lambda_{n,m}}\alpha_{m,1}(s) - \sqrt{1-\lambda_{n,m}}\alpha_{n,2}(t), \quad 0 \le s,t \le 1, \tag{37}$$

$$B_\lambda(s,t) = \sqrt{\lambda}B_1(s) - \sqrt{1-\lambda}B_2(t), \quad 0 \le s,t \le 1, \tag{38}$$

where $\lambda_{n,m} = \frac{n}{n+m} \to \lambda \in (0,1)$ and B_1, B_2 are independent Brownian bridges on $[0,1]$. Finally, we will write $\|\cdot\|_\infty$ for the sup norm in $[0,1]$ and $\omega_{m,1}(\delta), \omega_{n,2}(\delta)$ for the oscillation modulus of the empirical processes $\alpha_{m,1}$ and $\alpha_{n,2}$, respectively, namely,

$$\omega_{m,1}(\delta) = \sup_{0 \le t-s \le \delta} |\alpha_{m,1}(s) - \alpha_{m,1}(t)|$$

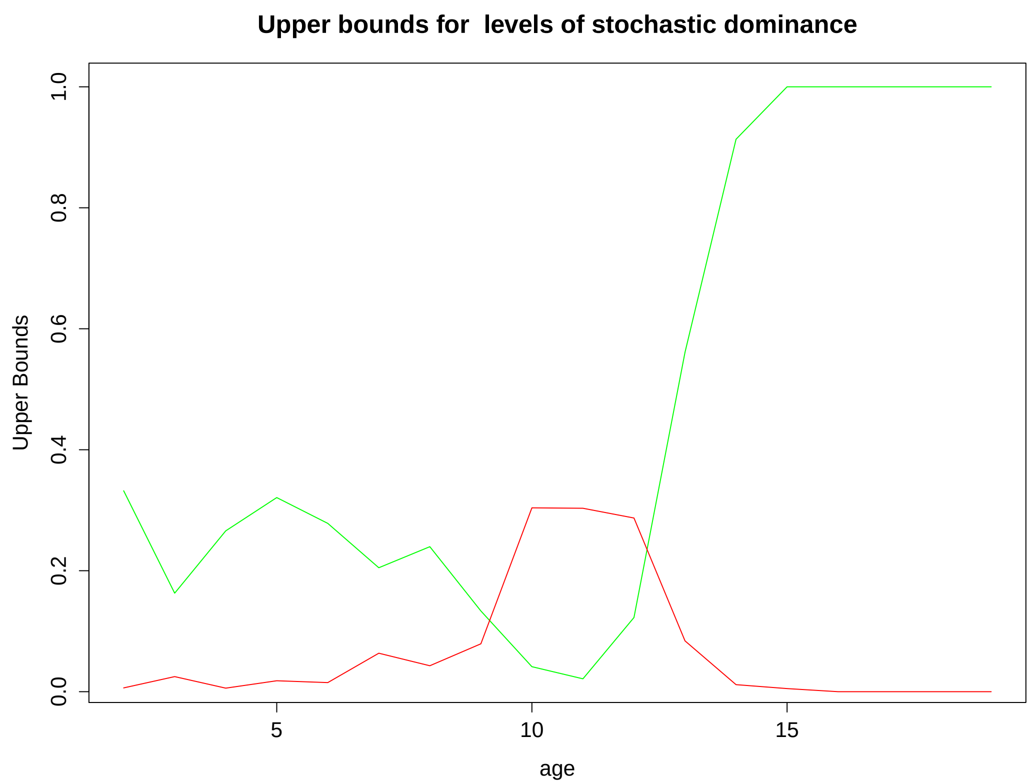

Figure 3: 95%-Confidence Upper bounds for the level of stochastic dominance of height of boys over girls (red) and of girls over boys (green) through the growth period $a = 2, ..., 19$.

and similarly for $\omega_{n,2}$.

The following estimates give the key to the asymptotic distributional behavior of the estimator $\pi(F_n, G_m)$.

Lemma 5.1 *If we denote* $\Delta_{n,m} = 2(m^{-1/2}\|\alpha_{m,1}\|_\infty + n^{-1/2}\|\alpha_{n,2}\|_\infty)$, $\tilde{\Gamma}_\delta(F,G) = \{x \in \mathbb{R} : G(x) - F(x) \geq \pi(F,G) - \delta\}$ *and*

$$R_{n,m} = \sqrt{\lambda_{n,m}}\omega_{m,1}(\Delta_{n,m}) + \sqrt{1-\lambda_{n,m}}\omega_{n,2}(\Delta_{n,m})$$

then

$$\begin{aligned} \sup_{x\in\Gamma(F,G)} \alpha_{n,m}(G(x), F(x)) &\leq \sqrt{\tfrac{mn}{m+n}}(\pi(F_n, G_m) - \pi(F,G)) \\ &\leq \sup_{x\in\tilde{\Gamma}_{\Delta_{n,m}}(F,G)} \alpha_{n,m}(G(x), F(x)) \end{aligned} \qquad (39)$$

and

$$\sqrt{\tfrac{mn}{m+n}}(\pi(F_n, G_m) - \pi(F,G)) \leq \sup_{t\in[\pi(F,G),1]} \alpha_{n,m}(t, t - \pi(F,G)) + R_{n,m}. \qquad (40)$$

Proof. We recall that $\Gamma(F,G) := \{x \in \bar{\mathbb{R}} : G(x)-F(x) = \pi(F,G)\}$. Hence, if $x \in \Gamma(F,G)$ then $G_m(x) - F_n(x) = (G_m(x) - G(x)) - (F_n(x) - F(x)) + \pi(F,G) = \sqrt{\frac{n+m}{nm}}\alpha_{n,m}(F(x) + \pi(F,G), F(x)) + \pi(F,G)$. From this obtain the lower bound in (39). Also, writing $G_m(x) - F_n(x) = (G_m(x) - G(x)) - (F_n(x) - F(x)) + (G(x) - F(x))$ we see that $G_m(x) - F_n(x) \leq G(x) - F(x) + m^{-1/2}\|\alpha_{m,1}\|_\infty + n^{-1/2}\|\alpha_{n,2}\|_\infty$ while for any $x \in \Gamma(F,G)$, $G_m(x) - F_n(x) \geq \pi(F,G) - m^{-1/2}\|\alpha_{m,1}\|_\infty - n^{-1/2}\|\alpha_{n,2}\|_\infty$. Therefore, for any x' outside $\tilde{\Gamma}_{\Delta_{n,m}}(F,G)$ and any $x \in \Gamma(F,G)$

$$G_m(x') - F_n(x') < \pi(F,G) - m^{-1/2}\|\alpha_{m,1}\|_\infty - n^{-1/2}\|\alpha_{n,2}\|_\infty \leq G_m(x) - F_n(x),$$

which means that $\pi(F_n, G_m) = \sup_{x \in \tilde{\Gamma}_{\Delta_{n,m}}(F,G)}(G_m(x) - F_n(x))$. As a consequence

$$\sqrt{\tfrac{mn}{m+n}}(\pi(F_n,G_m) - \pi(F,G)) \leq \sup_{x \in \tilde{\Gamma}_{\Delta_{n,m}}(F,G)} \sqrt{\lambda_{n,m}}\alpha_{m,1}(G(x)) - \sqrt{1-\lambda_{n,m}}\alpha_{n,2}(F(x)),$$

giving the upper bound in (39).

Consider now $x \in \tilde{\Gamma}_{\Delta_{n,m}}(F,G)$. If $G(x) \leq \pi(F,G)$ then $G(x) \geq \pi(F,G) - \Delta_{n,m}$ and $F(x) \leq \Delta_{n,m}$, from which we see that

$$\alpha_{n,m}(G(x), F(x)) \leq \alpha_{n,m}(\pi(F,G), 0) + R_{n,m}.$$

On the other hand, if $x \in \tilde{\Gamma}_{\Delta_{n,m}}(F,G)$ and $G(x) \geq \pi(F,G)$ then $G(x) - \pi(F,G) + \Delta_{n,m} \geq F(x) \geq G(x) - \pi(F,G) \geq 0$ and this entails

$$\alpha_{n,m}(G(x), F(x)) \leq \alpha_{n,m}(G(x), G(x) - \pi(F,G)) + R_{n,m}.$$

Combining the last two estimates we conclude (40). $\square$

Proof of Theorem 3.1. From (39) we see that the result will follow if we show that

$$\sup_{x \in \Gamma(F,G)} \alpha_{n,m}(G(x), F(x)) \xrightarrow{w} \sup_{x \in \Gamma(F,G)} B_\lambda(G(x), F(x)) \tag{41}$$

and

$$\sup_{x \in \tilde{\Gamma}_{\Delta_{m,n}}(F,G)} \alpha_{n,m}(G(x), F(x)) \xrightarrow{w} \sup_{x \in \Gamma(F,G)} B_\lambda(G(x), F(x)) \tag{42}$$

We can assume without loss of generality that $\alpha_{m,1}$ and $\alpha_{n,2}$ are defined on a rich enough probability space in which there are also independent Brownian bridges, for which we keep the notation B_1, B_2 such that $\|\alpha_{m,1} - B_1\|_\infty \to 0$ and $\|\alpha_{n,2} - B_2\|_\infty \to 0$ a.s. (see e.g. Theorem 1, p. 93 in Shorack and Wellner (1986)). Note that $\|\alpha_{n,m} - B_\lambda\|_\infty \to 0$ a.s., which implies that a.s.

$$\sup_{\delta \geq 0} \left| \sup_{x \in \tilde{\Gamma}_\delta(F,G)} \alpha_{n,m}(G(x), F(x)) - \sup_{x \in \tilde{\Gamma}_\delta(F,G)} B_\lambda(G(x), F(x)) \right| \leq \|\alpha_{n,m} - B_\lambda\|_\infty \to 0 \tag{43}$$

This (take $\delta = 0$), proves (41). We claim that

$$\sup_{x\in\tilde{\Gamma}_{\Delta_{n,m}}(F,G)} B_\lambda(G(x), F(x)) \to \sup_{x\in\Gamma(F,G)} B_\lambda(G(x), F(x)) \text{ a.s.} \tag{44}$$

In fact, by continuity, $\sup_{x\in\tilde{\Gamma}_{\Delta_{m,n}}(F,G)} B_\lambda(G(x), F(x)) = B_\lambda(G(x_n), F(x_n))$ for some $x_n \in \tilde{\Gamma}_{\Delta_{m,n}}(F, G)$ and by compactness, from any subsequence we can extract a further subsequence (that we keep denoting x_n) such that $x_n \to x_0$. Since $\Delta_{m,n} \to 0$ a.s., necessarily, $x_0 \in \Gamma(F, G)$ and $B_\lambda(G(x_n), F(x_n)) \to B_\lambda(G(x_0), F(x_0))$, which means that a.s.

$$\limsup_{n\to\infty} \sup_{x\in\tilde{\Gamma}_{\Delta_{m,n}}(F,G)} B_\lambda(G(x), F(x)) \leq \sup_{x\in\Gamma(F,G)} B_\lambda(G(x), F(x)).$$

Since, obviously, $\sup_{x\in\tilde{\Gamma}_{\Delta m,n}(F,G)} B_\lambda(G(x), F(x))) \geq \sup_{x\in\Gamma(F,G)} B_\lambda(G(x), F(x))$, we get (44). Using now (43) we conclude (42) and prove (14).

For the bootstrap result we note that

$$\sqrt{\tfrac{mn}{m+n}} \sup_{x\in\Gamma_{n,m}} ((G_m^*(x) - G_m(x)) - (F_n^*(x) - F_n(x)))) \stackrel{d}{=} \sup_{x\in\Gamma_{n,m}} \alpha'_{n,m}(G_m(x), F_n(x)),$$

where $\alpha'_{n,m}$ is an independent copy of $\alpha_{n,m}$ (hence, independent of the X_i's and Y_j's). We can argue as above and assume that there is an independent copy B_λ, that we denote B'_λ such that $\|\alpha_{n,m} - B'_\lambda\|_\infty \to 0$ a.s. Since a.s. $B'_\lambda(G_m(x), F_n(x)) \to B'_\lambda(G(x), F(x))$ we see that we simply have to prove that

$$V_{n,m}^* = \sup_{x\in\Gamma_{n,m}} B'_\lambda(F(x), G(x)) \to \sup_{x\in\Gamma(F,G)} B'_\lambda(F(x), G(x)) := V \quad \text{a.s.}$$

To check this, we note that, a.s., $G_m(x) - F_n(x) \to G(x) - F(x)$ uniformly in $x \in \bar{\mathbb{R}}$. Consider a sequence of points $x_n \in \Gamma_{n,m}$, that is, such that $G_m(x_n) - F_n(x_n) \geq \pi(F_n, G_m) - \delta_{n,m}$. ¿From any subsequence we can extract a further convergent subsequent for which, again, we keep the notation $x_n \to x_0$. Then, since $\delta_{n,m} \to 0$, $G(x_0) - F(x_0) = \lim_n(G_m(x_n) - F_n(x_n)) \geq \lim_n \pi(F_n, G_m) - \delta_{n,m} = \pi(F, G)$, that is, $x_0 \in \Gamma(F, G)$. This shows that $\limsup_{n\to\infty} V_{n,m}^* \leq V$. For the lower bound, we recall from Lemma 5.1 that for $x \in \Gamma(F, G)$, $G_m(x) - F_n(x) \geq \pi(F_n, G_m) - \Delta_{m,n}$ Now, the choice of δ_n and the law of iterated logarithm for the empirical process (see, e.g., Theorem 1, p. 504 in Shorack and Wellner (1986)) ensure that a.s.

$$\begin{aligned}\limsup_{n\to\infty} \frac{\Delta_{n,m}}{\delta_n} &= \limsup_{n\to\infty} \tfrac{2}{K}\left(\sqrt{\lambda_{n,m}}\tfrac{\|\alpha_{m,1}\|_\infty}{\sqrt{\log\log m}} + \sqrt{1-\lambda_{n,m}}\tfrac{\|\alpha_{n,2}\|_\infty}{\sqrt{\log\log n}}\right) \\ &= \tfrac{\sqrt{2}}{K}(\sqrt{\lambda} + \sqrt{1-\lambda}) \leq \tfrac{2}{K} < 1.\end{aligned}$$

Hence, eventually $\Delta_{n,m} < \delta_{n,m}$ and if $x \in \Gamma(F, G)$ then $G_m(x) - F_n(x) \geq \pi(F_n, G_m) - \delta_n$, that is, eventually $\Gamma(F, G) \subset \Gamma_{n,m}$. As a consequence we see that, with probability one, $\liminf_{n\to\infty} V_{n,m}^* \geq V$. This completes the proof. □

Next, we prove the results connected to the limiting distribution in Theorem 3.1.

Proof of Proposition 3.2. The upper bound for $K_\alpha(F,G,\lambda)$ follows from the obvious fact

$$\begin{aligned}\bar{B}(F,G,\lambda) &= \sup_{t\in T(F,G,\pi(F,G))}\left(\sqrt{\lambda}\ B_1(t)-\sqrt{1-\lambda}\ B_2(t-\pi(F,G))\right)\\ &\leq \sup_{t\in[\pi(F,G),1]}\left(\sqrt{\lambda}\ B_1(t)-\sqrt{1-\lambda}\ B_2(t-\pi(F,G))\right)=\bar{B}(\pi(F,G),\lambda).\end{aligned}$$

For the lower bound note that for every $t\in T(F,G,\pi(F,G))$ we have $\bar{B}(F,G,\lambda)\geq\sqrt{\lambda}\ B_1(t)-\sqrt{1-\lambda}\ B_2(t-\pi(F,G))$ and this last variable is centered, normally distributed with variance σ_t^2 and its α-quantile is, therefore, $\sigma_t\Phi^{-1}(\alpha)$. If $\alpha\geq\frac{1}{2}$ then the best lower bound of this kind is obtained for $\sigma_t=\bar{\sigma}(F,G,\pi(F,G))$, while for $\alpha<\frac{1}{2}$ we have $\Phi^{-1}(\alpha)<0$ and the largest upper bound is given by $\underline{\sigma}(F,G,\pi(F,G))\Phi^{-1}(\alpha)$. □

Proof of Proposition 3.3. We observe first that $\{\sqrt{\lambda}\ B_1(t)-\sqrt{1-\lambda}\ B_2(t-a)\}_{a\leq t\leq 1}$ has the same distribution as

$$\left\{\sqrt{1-a}B(\tfrac{t-a}{1-a})+\sqrt{\lambda a(1-a)}(1-\tfrac{t-a}{1-a})X+\sqrt{(1-\lambda)a(1-a)}\tfrac{t-a}{1-a}Y\right\}_{a\leq t\leq 1},$$

where B is another Brownian bridge and X and Y are independent standard normal r.v.'s, independent of B (just note that both processes are centered Gaussian with the same covariance function, namely, $\lambda s(1-t)+(1-\lambda)(s-a)(1-t+a)$ for $a\leq s\leq t\leq 1$). This implies that

$$\bar{B}(a,\lambda)\overset{d}{=}\sqrt{1-a}\sup_{0\leq s\leq 1}\left(B(s)+\sqrt{\lambda a}(1-s)X+\sqrt{(1-\lambda)a}sY\right)\tag{45}$$

with B,X and Y as above. From this point, we focus, for simplicity, on the case $\lambda=\frac{1}{2}$, the general case following with straighforward but tedious, changes from this. Using the well-known fact that

$$P\left(\sup_{0\leq t\leq 1}(B(t)-\alpha(1-t)-\beta t)>0\right)=\begin{cases}e^{-2\alpha\beta} & \text{if } \alpha>0,\beta>0\\ 1 & \text{otherwise}\end{cases},\tag{46}$$

see, e.g., Hájek et al. (1999), p. 219, we see that

$$\begin{aligned}P(\bar{B}(a,\lambda)>u\sqrt{1-a}) &= P\left(\sup_{0\leq s\leq 1}\left(B(s)+\sqrt{a/2}(1-s)X+\sqrt{a/2}sY\right)>u\right)\\ &= 1-\Phi(\tfrac{\sqrt{2}u}{\sqrt{a}})^2+\int_{x\leq\frac{\sqrt{2}u}{\sqrt{a}},y\leq\frac{\sqrt{2}u}{\sqrt{a}}}e^{-2(u-\sqrt{\frac{a}{2}}x)(u-\sqrt{\frac{a}{2}}y)}\frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}dxdy\\ &= 1-\Phi(\tfrac{\sqrt{2}u}{\sqrt{a}})^2+e^{-\frac{2u^2}{1+a}}\int_{-\infty}^{\frac{\sqrt{2}u}{\sqrt{a}}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1-a^2}{2}(x-\frac{\sqrt{2a}u}{1+a})^2}\Phi\left(\frac{\sqrt{2}u(1-a)}{\sqrt{a}}+ax\right)dx\end{aligned}$$

and conclude (a). To prove (b), we write

$$U = \sup_{0\le t\le 1} (B(t) + \alpha(1-t) + \beta t).$$

and note from (46) that $P(U > u) = e^{-2(u-\alpha)(u-\beta)}$ for $u \ge \max(\alpha, \beta)$ and $P(U > u) = 1$ otherwise. Hence, U has density $2(2u - (\alpha + \beta))e^{-2(u-\alpha)(u-\beta)}$, $u \ge \max(\alpha, \beta)$ and if we write $M(t) = E(e^{tU})$ for the moment generating function of U, then with the change of variable $u = v + \frac{\alpha+\beta}{2}$ we obtain

$$\begin{aligned}
M(t) &= \int_{\alpha\vee\beta}^{\infty} 2(2u-(\alpha+\beta))e^{-2(u-\alpha)(u-\beta)}e^{tu}du \\
&= e^{\frac{(\alpha-\beta)^2+t(\alpha+\beta)+\frac{t^2}{4}}{2}} \int_{\frac{|\alpha-\beta|}{2}}^{\infty} 4ve^{-2(v-\frac{t}{4})^2}dv \\
&= e^{(\alpha\vee\beta)t}e^{\frac{1}{2}(|\alpha-\beta|-\frac{t}{2})^2}\left[\int_{\frac{|\alpha-\beta|}{2}}^{\infty} 4(v-\tfrac{t}{4})e^{-2(v-\frac{t}{4})^2}dv + t\int_{\frac{|\alpha-\beta|}{2}}^{\infty} e^{-2(v-\frac{t}{4})^2}dv\right] \\
&= e^{(\alpha\vee\beta)t}\left[1 + t\sqrt{\tfrac{\pi}{2}}e^{\frac{1}{2}(|\alpha-\beta|-\frac{t}{2})^2}(1-\Phi(|\alpha-\beta|-\tfrac{t}{2}))\right] \\
&= e^{(\alpha\vee\beta)t}\left[1 + \frac{t}{2}\frac{(1-\Phi(|\alpha-\beta|-\frac{t}{2}))}{\varphi(|\alpha-\beta|-\frac{t}{2})}\right].
\end{aligned}$$

Differentation in this last expression yields now

$$E(U) = M'(0) = \alpha\vee\beta + \tfrac{1}{2}\tfrac{(1-\Phi(|\alpha-\beta|))}{\varphi(|\alpha-\beta|)},$$

$$E(U^2) = M''(0) = (\alpha\vee\beta)^2 + \tfrac{1}{2} + (\alpha+\beta)\tfrac{1}{2}\tfrac{(1-\Phi(|\alpha-\beta|))}{\varphi(|\alpha-\beta|)},$$

Now, taking $\alpha = \sqrt{a/2}X$, $\beta = \sqrt{a/2}Y$ and taking expectations in the resulting expression we obtain

$$E(\bar{B}(a, \tfrac{1}{2})) = \sqrt{\tfrac{a(1-a)}{2}}E(\max(X,Y)) + \sqrt{\tfrac{\pi(1-a)}{2}}E(e^{\frac{aZ^2}{2}}(1-\Phi(\sqrt{a}|Z|))), \tag{47}$$

where $Z = (X-Y)/\sqrt{2}$ is standard normal. Observe now that

$$\begin{aligned}
E(e^{\frac{aZ^2}{2}}(1-\Phi(\sqrt{a}|Z|))) &= \tfrac{2}{\sqrt{2\pi}}\int_0^{\infty} e^{-\frac{(1-a)z^2}{2}}(1-\Phi(\sqrt{a}z))dz \\
&= \tfrac{2}{2\pi}\int_{(0<\sqrt{a}z<y)} e^{-\frac{(1-a)z^2+y^2}{2}}dzdy = \tfrac{1}{\pi\sqrt{1-a}}\int_{(0<\sqrt{a/(1-a)}x<y)} e^{-\frac{x^2+y^2}{2}}dxdy \\
&= \tfrac{1}{\pi\sqrt{1-a}}\int_{(0<r<\infty, \operatorname{atan}\left(\sqrt{\frac{a}{1-a}}\right)<\theta<\frac{\pi}{2})} re^{-\frac{r^2}{2}}drd\theta = \tfrac{1}{\pi\sqrt{1-a}}\left[\tfrac{\pi}{2} - \operatorname{atan}\left(\sqrt{\tfrac{a}{1-a}}\right)\right].
\end{aligned}$$

Plugging this into (47) and taking into account that $E(\max(X,Y)) = \frac{1}{\sqrt{\pi}}$ we obtain the conclusion about $E(\bar{B}(a, \frac{1}{2}))$. A similar computation yields $E(\bar{B}(a, \frac{1}{2})) = \frac{1-a^2}{2}$ and completes the proof. $\square$

Next, we prove the result about the level and power of the test for essential stochastic order.

Proof of Proposition 3.4. We assume for simplicity $m = n$. The general case can be handled with straighforward changes. A simple computation shows that $\pi(F_0, G_0) = \pi_0$ and $\Gamma(F_0, G_0) = \{\frac{1+\pi_0}{2}\}$ and, using Theorem 3.1, that $\sqrt{\frac{n}{2}}(\pi_{n,n} - \pi_0) \xrightarrow{w} N(0, \bar{\sigma}^2_{\pi_0})$. Hence,

$$\begin{aligned} \liminf_{n\to\infty} \sup_{(F,G)\in H_0} & \mathbb{P}_{F,G}\left[\sqrt{\tfrac{n}{2}}(\pi_{n,n} - \pi_0) < \bar{\sigma}_{\pi_0}\Phi^{-1}(\alpha)\right] \\ & \geq \lim_{n\to\infty} \mathbb{P}_{F_0,G_0}\left[\sqrt{\tfrac{n}{2}}(\pi_{n,n} - \pi_0) < \bar{\sigma}_{\pi_0}\Phi^{-1}(\alpha)\right] = \alpha. \end{aligned}$$

For the upper bound we recall from (39) that $\sqrt{\frac{n}{2}}(\pi_{n,n} - \pi(F,G)) \geq \alpha_{n,n}(G(x), G(x) - \pi(F,G))$ for every $x \in \Gamma(F,G)$. As a consequence, for $(F,G) \in H_0$ and $x \in \Gamma(F,G)$,

$$\begin{aligned} & \mathbb{P}_{F,G}\left[\sqrt{\tfrac{n}{2}}(\pi_{n,n} - \pi_0) \leq \bar{\sigma}_{\pi_0}\Phi^{-1}(\alpha)\right] \\ & \quad \leq P\left[\alpha_{n,n}(G(x), G(x) - \pi(F,G)) \leq \bar{\sigma}_{\pi_0}\Phi^{-1}(\alpha) - \sqrt{\tfrac{n}{2}}(\pi(F,G) - \pi_0)\right]. \qquad (48) \end{aligned}$$

We observe that, for any $x \in \Gamma(F,G)$, $\alpha_{n,n}(G(x), G(x) - \pi(F,G))$ is a sum of n i.i.d. centered random variables with variance $\sigma^2(x) = \frac{1}{2}G(x)(1-G(x)) + \frac{1}{2}(G(x) - \pi(F,G))(1 - (G(x) - \pi(F,G)))$ and third absolute moment smaller than $2^{3/2}$. From the Berry-Esseen inequality (see, e.g., Theorem 1, p. 848 in Shorack and Wellner (1986)) we see that for some universal constant $C > 0$

$$\mathbb{P}_{F,G}\left[\sqrt{\tfrac{n}{2}}(\pi_{n,n} - \pi_0) \leq \bar{\sigma}_{\pi_0}\Phi^{-1}(\alpha)\right] \leq \Phi\left(\frac{\bar{\sigma}_{\pi_0}\Phi^{-1}(\alpha) - \sqrt{\frac{n}{2}}(\pi(F,G) - \pi_0)}{\sigma(x)}\right) + \frac{2^{3/2}C}{\sigma(x)^{3/2}\sqrt{n}}.$$

The computations leading to the expressions for $\bar{\sigma}^2_\pi$ and $\underline{\sigma}^2_\pi$ show that $\frac{1}{2}\pi(F,G)(1 - \pi(F,G)) \leq \sigma(x) \leq \frac{1}{4}(1 - \pi^2(F,G)) \leq \frac{1}{4}(1 - \pi_0) = \bar{\sigma}^2_{\pi_0}$ for $(F,G) \in H_0$ and $x \in \Gamma(F,G)$. This and the fact that $\bar{\sigma}_{\pi_0}\Phi^{-1}(\alpha) - \sqrt{\frac{n}{2}}(\pi(F,G) - \pi_0) \leq 0$ yield that for every $(F,G) \in H_0$

$$\begin{aligned} \mathbb{P}_{F,G}\left[\sqrt{\tfrac{n}{2}}(\pi_{n,n} - \pi_0) \leq \bar{\sigma}_{\pi_0}\Phi^{-1}(\alpha)\right] & \leq \Phi\left(\Phi^{-1}(\alpha) - \tfrac{\sqrt{n}}{2\bar{\sigma}_{\pi_0}}(\pi(F,G) - \pi_0)\right) \\ & \quad + \frac{8C}{(\pi(F,G)(1 - \pi(F,G)))^{3/2}\sqrt{n}} \\ & \leq \alpha + \frac{8C}{(\pi(F,G)(1 - \pi(F,G)))^{3/2}\sqrt{n}}. \end{aligned}$$

On the other hand, from (48) and Hoeffding's inequality we see that

$$\begin{aligned} \mathbb{P}_{F,G}\left[\sqrt{\tfrac{n}{2}}(\pi_{n,n} - \pi_0) \leq \bar{\sigma}_{\pi_0}\Phi^{-1}(\alpha)\right] & \leq e^{-2(\bar{\sigma}_{\pi_0}\Phi^{-1}(\alpha) - \sqrt{\frac{n}{2}}(\pi(F,G) - \pi_0))^2} \\ & \leq e^{-n(\pi(F,G) - \pi_0)^2}. \qquad (49) \end{aligned}$$

This, in particular, yields (24). Now fix $\delta > 0$ small enough to ensure that $\pi_0 + \delta < 1$ and $\pi(1-\pi) \geq \frac{1}{2^{2/3}}\pi_0(1-\pi_0)$ if $\pi_0 + \delta \geq \pi \geq \pi_0$. Then

$$\sup_{(F,G)\in H_0} \mathbb{P}_{F,G}\left[\sqrt{\tfrac{n}{2}}(\pi_{n,n} - \pi_0) \leq \bar{\sigma}_{\pi_0}\Phi^{-1}(\alpha)\right] \leq \alpha + \frac{16C}{(\pi_0(1-\pi_0))^{3/2}\sqrt{n}} + e^{-\frac{n}{2}\delta} \to \alpha$$

and this proves (23).

Finally, for the proof of (25) we simply note that $\pi_{n,n} - \pi(F,G) \leq n^{-1/2}(\|\alpha_{n,1}\|_\infty + \|\alpha_{n,2}\|_\infty)$ and, therefore,

$$\begin{aligned}
&\mathbb{P}_{F,G}\left[\sqrt{\tfrac{n}{2}}(\pi_{n,n} - \pi_0) > \bar{\sigma}_{\pi_0}\Phi^{-1}(\alpha)\right] \\
&\quad = \mathbb{P}_{F,G}\left[\sqrt{\tfrac{n}{2}}(\pi_{n,n} - \pi(F,G)) > \bar{\sigma}_{\pi_0}\Phi^{-1}(\alpha) + \sqrt{\tfrac{n}{2}}(\pi_0 - \pi(F,G))\right] \\
&\quad \leq P(\|\alpha_{n,1}\|_\infty > K/\sqrt{2}) + P(\|\alpha_{n,2}\|_\infty > K/\sqrt{2}),
\end{aligned}$$

where $K = \bar{\sigma}_{\pi_0}\Phi^{-1}(\alpha) + \sqrt{\frac{n}{2}}(\pi_0 - \pi(F,G))$. An application of the Dvoretzky-Kiefer-Wolfowitz inequality, see Massart (1990), yields

$$\mathbb{P}_{F,G}\left[\sqrt{\tfrac{n}{2}}(\pi_{n,n} - \pi_0) > \bar{\sigma}_{\pi_0}\Phi^{-1}(\alpha)\right] \leq 2e^{-K^2}$$

and completes the proof. □

Proof of Proposition 3.5. We keep the notation of the proof of Theorem 3.1 with $G_m(x) = G(x) + \frac{1}{\sqrt{m}}\alpha_{m,1}(G(x))$ and $F_n(x) = F(x) + \frac{1}{\sqrt{n}}\alpha_{n,2}(F(x))$ for independent uniform empirical processes $\alpha_{m,1}$, $\alpha_{n,2}$ that we assume, without loss of generality, to be defined on a rich enough probability space in which there are independent Brownian bridges, $B_{m,1}$, $B_{n,2}$, satisfying

$$P\left[\|\alpha_{m,1} - B_{m,1}\|_\infty > m^{-1/2}(x + 12\log m)\right] \leq 2e^{-x/6}, \quad x > 0 \tag{50}$$

and similarly for $\alpha_{n,2}$ and $B_{n,2}$ (see, e.g., Csörgo and Horváth (1989), p. 114). In particular, we have that $E(\|\alpha_{m,1} - B_{m,1}\|_\infty) \leq \frac{12(1+\log m)}{\sqrt{m}}$. We define $\tilde{G}_m(x) = G(x) + \frac{1}{\sqrt{m}}B_{m,1}(G(x))$, $\tilde{F}_n(x) = F(x) + \frac{1}{\sqrt{n}}B_{n,2}(F(x))$ and $\pi(\tilde{F}_n, \tilde{G}_m) = \sup_{x\in\mathbb{R}}(\tilde{G}_m(x) - \tilde{F}_n(x))$. From (50) we obtain that

$$\begin{aligned}
&\sqrt{\tfrac{nm}{n+m}}E|\pi(F_n, G_m) - \pi(\tilde{F}_n, \tilde{G}_m)| \\
&\quad \leq \sqrt{\lambda_{n,m}}E(\|\alpha_{m,1} - B_{m,1}\|_\infty) + \sqrt{1-\lambda_{m,n}}E(\|\alpha_{n,2} - B_{n,2}\|_\infty) \to 0,
\end{aligned} \tag{51}$$

as $n, m \to \infty$. We can write, as well, $G_m^*(x) = G_m(x) + \frac{1}{\sqrt{m}}\alpha'_{m,2}(G_m(x))$, $F_n^*(x) = F_n(x) + \frac{1}{\sqrt{n}}\alpha'_{n,2}(F_n(x))$ with $\alpha'_{m,1}$, $\alpha'_{n,2}$ independent uniform empirical processes, independent of $\alpha_{m,1}$ and $\alpha_{n,2}$ and $B'_{m,1}$, $B'_{n,2}$ for Brownian bridges related to $\alpha'_{m,1}$, $\alpha'_{n,2}$ as in (50). Now, we define $\tilde{G}^*_m(x) = \tilde{G}_m(x) + \frac{1}{\sqrt{m}}\alpha'_{m,1}(G_m(x))$, $\tilde{F}^*_n(x) = \tilde{F}_n(x) + \frac{1}{\sqrt{n}}\alpha'_{n,2}(\tilde{F}_n(x))$, $\pi(\tilde{F}^*_n, \tilde{G}^*_m) = \sup_{x\in\mathbb{R}}(\tilde{G}^*_m(x) - \tilde{F}^*_n(x))$. We claim that

$$\sqrt{\tfrac{nm}{n+m}}E|\pi(F^*_n, G^*_m) - \pi(\tilde{F}^*_n, \tilde{G}^*_m)| \to 0. \tag{52}$$

In fact, after estimate (51) it suffices to show that

$$E\left(\sup_x |\alpha'_{m,1}(G_m(x)) - \alpha'_{m,1}(\tilde{G}_m(x))|\right) \to 0 \tag{53}$$

and the same for $\alpha'_{n,2}$. But we have that $E\sup_{0\leq t-s\leq a}|B'_{m,1}(t)-B'_{m,1}(s)| \leq 7\sqrt{2a\log(1/a)}$, for $a \in (0, \frac{1}{2}]$ (see Theorem 3, p. 538, in Shorack and Wellner (1986)). Hence, conditioning on $G_m, \tilde{G}_m$ and taking $\nu \in (0, \frac{1}{2})$ we obtain from (50) that for some positive constant, K,

$$E\left(\sup_x |\alpha'_{m,1}(G_m(x)) - \alpha'_{m,1}(\tilde{G}_m(x))|\right) \leq K\left(\tfrac{\log m}{\sqrt{m}} + \tfrac{1}{m^{\nu/2}}E(\|\alpha_{m,1} - B_{m,1}\|^s_\infty)\right) \to 0.$$

Now, from (51) and (52) we see that the result will follow if we prove that

$$\sqrt{\tfrac{nm}{n+m}}E|\pi(\tilde{F}^*_n, \tilde{G}^*_m) - \pi(\tilde{F}_n, \tilde{G}_m)| \to 0. \tag{54}$$

As in the proof of Theorem 3.1 we see that

$$\sup_{x\in\Gamma(\tilde{F}_n,\tilde{G}_m)} \alpha'_{n,m}(x) \leq \sqrt{\tfrac{nm}{n+m}}(\pi(\tilde{F}^*_n, \tilde{G}^*_m) - \pi(\tilde{F}_n, \tilde{G}_m)) \leq \sup_{x\in\bar{\Gamma}(\tilde{F}_n,\tilde{G}_m)} \alpha'_{n,m}(x),$$

where $\alpha'_{n,m}(x) = \sqrt{\lambda_{n,m}}\alpha'_{m,1}(\tilde{G}_m(x)) + \sqrt{1-\lambda_{n,m}}\alpha'_{n,2}(\tilde{F}_n(x))$, $\Gamma(\tilde{F}_n, \tilde{G}_m) = \{x : \tilde{G}_m(x) - \tilde{F}_n(x) = \pi(\tilde{F}_n, \tilde{G}_m)\}$ and $\bar{\Gamma}(\tilde{F}_n, \tilde{G}_m) = \{x : \tilde{G}_m(x) - \tilde{F}_n(x) \geq \pi(\tilde{F}_n, \tilde{G}_m) - 2(\|\alpha'_{m,1}\|_\infty/\sqrt{m} + \|\alpha'_{n,2}\|_\infty/\sqrt{n})$. We can mimic the argument in Theorem 3.1 to show that

$$\sup_{x\in\bar{\Gamma}(\tilde{F}_n,\tilde{G}_m)} \alpha'_{n,m}(x) - \sup_{x\in\Gamma(\tilde{F}_n,\tilde{G}_m)} \alpha'_{n,m}(x) \to 0$$

in L_1. Finally, to show that

$$E\Big(\sup_{x\in\Gamma(\tilde{F}_n,\tilde{G}_m)} \alpha'_{n,m}(x)\Big) \to 0,$$

observe that $\tilde{G}_m(x) - \tilde{F}_n(x)$ is a Gaussian process with continuous sample paths whose increments have nonzero variance. As a consequence (see Lemma 2.6 in Kim and Pollard (1990)), with probability one, $\Gamma(\tilde{F}_n, \tilde{G}_m)$ consists of just one point, say $x_{n,m}$, which depends on $\tilde{F}_n$ and $\tilde{G}_m$. Conditionally given $\tilde{F}_n$ and $\tilde{G}_m$, $\sup_{x\in\Gamma(\tilde{F}_n,\tilde{G}_m)} \alpha'_{n,m}(x) = \alpha'_{n,m}(x_{n,m})$ is a centered random variable. But taking expectations we see that, in fact,

$$E\Big(\sup_{x\in\Gamma(\tilde{F}_n,\tilde{G}_m)} \alpha'_{n,m}(x)\Big) = 0.$$

This completes the proof. $\square$

Proof of Proposition 3.6. We only deal with (31) since (32) follows from the Dvoretzky-Kiefer-Wolfowitz inequality and (33) from Hoeffding's inequality reproducing almost verbatim the arguments in Proposition 3.4. Also, for ease of notation, we consider the case $m = n$. For the pair F_0, G_0 we have $\pi(F_0, G_0) = \pi_0$ and $T(F_0, G_0, \pi_0) = [\pi_0, 1]$. This and Theorem 3.1 imply that

$$\begin{aligned}\liminf_{n\to\infty} \sup_{(F,G)\in H_0} & \mathbb{P}_{F,G}\left[\sqrt{\tfrac{n}{2}}(\pi_{n,n} - \pi_0) > K_{1-\alpha}(\pi_0, \tfrac{1}{2})\right] \\ & \geq \lim_{n\to\infty} \mathbb{P}_{F_0,G_0}\left[\sqrt{\tfrac{n}{2}}(\pi_{n,n} - \pi_0) > K_{1-\alpha}(\pi_0, \tfrac{1}{2})\right] = \alpha.\end{aligned}$$

To complete the proof of (31) assume, without loss of generality, that, as in the proof of Proposition 3.5, $\alpha_{n,1}$ and $\alpha_{n,2}$ are defined on a rich enough probability space together with Brownian bridges $B_{n,1}, B_{n,2}$ satisfying (50). In particular, if $B_n(s,t) = \frac{1}{2}B_{n,1}(s) + \frac{1}{2}B_{n,2}(t)$, then there are universal constants $c_1, c_2 > 0$ such that

$$P(\|\alpha_{n,n} - B_n\| \geq c_1 \tfrac{\log n}{\sqrt{n}}) \leq \tfrac{c_2}{n^2}. \tag{55}$$

Recall from Lemma 5.1 that

$$\sqrt{\tfrac{n}{2}}(\pi_{n,n} - \pi_0) \leq \sup_{t\in[\pi(F,G),1]} \alpha_{n,n}(t, t - \pi(F,G)) + R_{n,n} \tag{56}$$

with $R_{n,n} = \frac{1}{\sqrt{2}}(\omega_{n,1}(\Delta_{n,n}) + \omega_{n,2}(\Delta_{n,n}))$. We saw in the proof of Theorem 3.1 that $\limsup_{n\to\infty} \frac{\Delta_{n,n}}{\delta_n} = \frac{2}{K} < 1$ a.s. if $\delta = K\sqrt{\frac{2}{n}\log\log n}$ and $K > 2$. This implies that a.s., eventually $\omega_{n,1}(\Delta_{n,n}) \leq \omega_{n,1}(\delta_n)$. From Stute's results on the oscillation of the empirical process (see, e.g. Theorem 1, p. 542 in Shorack and Wellner (1986)) we have that a.s.

$$\lim_{n\to\infty} \frac{\sqrt{n}\omega_{n,1}(\delta_n)}{\sqrt{K\sqrt{2}\log n \log\log n}} = 1.$$

Consequently, $\frac{\sqrt{n}\omega_{n,1}(\Delta_{n,n})}{\log n} \to 0$ a.s. and the same happens for $\omega_{n,2}(\Delta_{n,n})$. Hence, $\frac{\sqrt{n}R_{n,n}}{\log n} \to 0$ a.s. and, in particular

$$P\left[R_{n,n} > \tfrac{\log n}{\sqrt{n}}\right] \to 0. \tag{57}$$

Now, combining (56), (55) and (57) we obtain that

$$\begin{aligned}&\mathbb{P}_{F,G}\left[\sqrt{\tfrac{n}{2}}(\pi_{n,n} - \pi_0) > K_{1-\alpha}(\pi_0, \tfrac{1}{2})\right] \\ &\quad \leq P\left[\sup_{t\in[\pi(F,G),1]} \alpha_{n,n}(t, t-\pi(F,G)) + R_{n,n} > K_{1-\alpha}(\pi_0, \tfrac{1}{2}) + \sqrt{\tfrac{n}{2}}(\pi_0 - \pi(F,G))\right] \\ &\quad \leq P\left[\sup_{t\in[\pi(F,G),1]} B_n(t, t-\pi(F,G)) > K_{1-\alpha}(\pi_0, \tfrac{1}{2}) - \tfrac{(c_1+1)\log n}{\sqrt{n}} + \sqrt{\tfrac{n}{2}}(\pi_0 - \pi(F,G))\right] \\ &\qquad + P\left[R_{n,n} > \tfrac{\log n}{\sqrt{n}}\right] + \tfrac{c_2}{n^2}.\end{aligned}$$

This shows that if suffices to prove that

$$\limsup_{n\to\infty} \sup_{a\le\pi_0} P\left[\bar{B}(a,\tfrac{1}{2}) > K_{1-\alpha}(\pi_0,\tfrac{1}{2}) + \sqrt{\tfrac{n}{2}}(\pi_0 - a) - r_n\right] \le \alpha \tag{58}$$

if $r_n \searrow 0$. To check this we note that the distribution function of $P(\bar{B}(a,\frac{1}{2}) \le x)$ depends continuously on (a,x) (this follows easily from Proposition 3.3 (a), for instance). Hence, given $\varepsilon > 0$ we can find $\pi_1 < \pi_0$ and $\delta > 0$ such that $P\Big(\bar{B}(a,\frac{1}{2}) > K_{1-\alpha}(\pi_0,\frac{1}{2}) - r\Big) \le \alpha + \varepsilon$ if $\pi_1 \le a \le \pi_0$ and $0 \le r \le \delta$. But then, taking n large enough to ensure that $r_n \le \delta$ we have

$$\begin{aligned}\sup_{\pi_1\le a\le\pi_0} & P\left[\bar{B}(a,\tfrac{1}{2}) > K_{1-\alpha}(\pi_0,\tfrac{1}{2}) + \sqrt{\tfrac{n}{2}}(\pi_0 - a) - r_n\right] \\ & \le \sup_{\pi_1\le a\le\pi_0} P\left[\bar{B}(a,\tfrac{1}{2}) > K_{1-\alpha}(\pi_0,\tfrac{1}{2}) - \delta\right] \le \alpha + \epsilon,\end{aligned}$$

while

$$\begin{aligned}\sup_{a\le\pi_1} P & \left[\bar{B}(a,\tfrac{1}{2}) > K_{1-\alpha}(\pi_0,\tfrac{1}{2}) + \sqrt{\tfrac{n}{2}}(\pi_0 - a) - r_n\right] \\ & \le \sup_{a\le\pi_1} P\left[\bar{B}(a,\tfrac{1}{2}) > \sqrt{\tfrac{n}{2}}(\pi_0 - \pi_1)\right] \\ & \le P\left[\|B_1\|_\infty + \|B_2\|_\infty > \sqrt{n}(\pi_0 - \pi_1)\right] \\ & \le 2e^{-\frac{n}{2}(\pi_0-\pi_1)^2}.\end{aligned}$$

The last two estimates complete the proof. □

References

Álvarez-Esteban, P.C.; del Barrio, E.; Cuesta-Albertos, J.A. and Matrán, C. (2008). Trimmed comparison of distributions. *J. Amer. Statist. Assoc.* 103, No. 482, 697–704.

Álvarez-Esteban, P.C.; del Barrio, E.; Cuesta-Albertos, J.A. and Matrán, C. (2011). Uniqueness and approximate computation of optimal incomplete transportation plans. *Annales de lInstitut Henri Poincaré - Probabilités et Statistiques*, 47, No. 2, 358–375.

Álvarez-Esteban, P.C.; del Barrio, E.; Cuesta-Albertos, J.A. and Matrán, C. (2012). Similarity of samples and trimming. *Bernoulli*, 18, 606–634.

Anderson, G. (1996). Nonparametric Tests for Stochastic Dominance. *Econometrica* 64, 1183–1193.

Arcones, M. A., Kvam, P. H., and Samaniego, F. J. (2002). Nonparametric Estimation of a Distribution Subject to a Stochastic Precedence Constraint. *J. Amer. Statist. Assoc.* 97, No. 457, 170–182.

Barrett, G.F., and Donald, S.G., (2003). Consistent tests for stochastic dominance. *Econometrica* 71, 71–104

Barron, A.R. (1989). Uniformly Powerful Goodness of Fit Tests. *Ann. Statist.,* **17**, 107–124.

Berger, R.L. (1988). A Nonparametric, Intersection-Union Test for Stochastic Order. In *Statistical Decision Theory and Related Topics IV. Volume 2.* (eds. Gupta, S. S., and Berger, J. O.). Springer-Verlag, New York.

Csörgo, M. and Horváth, L. (1989). *Weighted Approximations in Probability and Statistics.* Wiley.

Cuesta-Albertos, J. A., Gordaliza, A. and Matrán, C. (1997). Trimmed k-means: an attempt to robustify quantizers. *Ann. Statist.* **25**, 553–576.

Davidson, R., and Duclos, J.-Y. (2000). Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality. *Econometrica* 68, 1435–1464.

Davidson, R., and Duclos, J.-Y. (2013). Testing for Restricted Stochastic Dominance. *Econometric Reviews* **32**, 84–125.

Durbin, J. (1973). Distribution Theory for Tests Based on Sample Distribution Function. SIAM. CBMS-NSF Regional Conference Series in Applied Mathematics.

García-Escudero, L. A., Gordaliza, A., Matrán, C. and Mayo-Iscar, A. (2008). A general trimming approach to robust cluster analysis. *Ann. Statist.,* **36**, 1324–1345.

Gordaliza, A. (1991). Best approximations to random variables based on trimming procedures. *J. Approx. Theor.,* **64**, 162–180.

Hájek, J., Šidák, Z. and Sen, P.K. (1999). *Theory of Rank Tests.* Academic Press.

Hodges, J.L. and Lehmann, E.L. (1954). Testing the approximate validity of statistical hypotheses. *J. R. Statist. Soc. B* **16**, 261–268.

Kim, J. and Pollard, D. (1990). Cube root asymptotics. *Ann. Statist.,* **18**, 191–219.

Lehmann, E. L. (1955). Ordered families of distributions. *The Annals of Mathematical Statistics* 26, 399–419.

Lindsay, B.G., and Liu,J. (2009). Model Assessment Tools for a Model False World. *Statistical Science.* **24**, No 3, 303–318.

Linton, O., Maasoumi, E., and Whang, Y.-J. (2005). Consistent Testing for Stochastic Dominance under General Sampling Schemes. *Review of Economic Studies* 72, 735–765.

Linton, O., Song, K., and Whang, Y.-J. (2010). An improved bootstrap test for stochastic dominance. *Journal of Econometrics* 154, 186–202.

Liu, J., and Lindsay, B.G. (2009). Building and using semiparametric tolerance regions for parametric multinomial models. *Ann. Statist.,* **37**, 3644–3659.

Maronna, R. (2005). Principal components and orthogonal regression based on robust scales. *Technometrics* **47**, 264–273.

Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality, *Ann. Probab.,* **18**, 1269–1283.

McFadden, D. (1989). Testing for Stochastic Dominance. In *Studies in the Economics of Uncertainty: In Honor of Josef Hadar,* ed. by T. B. Fomby and T. K. Seo. New York. Springer.

Mosler, K. (1995). Testing Whether Two Distributions are Stochastically Ordered or Not. In *Grundlagen der Statistik und ihre Anwendungen: Festschrift für Kurt Weichselberger,* eds. by H. Rinne, B. Rüger, and H. Strecker. Heidelberg. Springer.

Müller, A., and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks.* Chichester, England. Wiley.

Raghavachari, M. (1973). Limiting distributions of the Kolmogorov-Smirnov type statistics under the alternative. *Ann. Statist.,* **1**, 67–73.

Rousseeuw, P. (1985). Multivariate estimation with high breakdown point. In W. Grossmann, G. Pflug, I. Vincze, and W. Werz (Eds.), *Mathematical Statistics and Applications, Volume B*, Reidel, Dordrecht.

Rudas, T., Clogg, C.C., and Lindsay, B. G. (1994). A New Index of Fit Based on Mixture Methods for the Analysis of Contingency Tables. *J. R. Statist. Soc. B* 56, No 4, 623–639.

Shaked, M., and Shanthikumar, J.G. (2007). *Stochastic Orders.* New York. Springer.

Schmid, F., and Trede, M. (1996 a). Testing for first order stochastic dominance in either direction. *Computational Statistics* **11**, 165–173.

Schmid, F., and Trede, M. (1996 b). Testing for first-order stochastic dominance: a new distribution-free test. *The Statistician* **45**, No. 3, 371–380.

Shorack, G.R. and Wellner, J.A. (1986). *Emprical Processes with Applications to Statistics.* Wiley.

Xi, L., and Lindsay, B.G. (1987). A note on calculating the π^* index of fit for the analysis of contingency tables. *Sociological Methods and Research* **25**, 248–259.