ESAIM: PS 23 (2019) 697–738
https://doi.org/10.1051/ps/2018022

ESAIM: Probability and Statistics
www.esaim-ps.org

INTERMEDIATE EFFICIENCY IN NONPARAMETRIC TESTING PROBLEMS WITH AN APPLICATION TO SOME WEIGHTED STATISTICS*

Tadeusz Inglot[1], Teresa Ledwina[2] and Bogdan Ćmiel[3,**]

Abstract. The basic motivation and primary goal of this paper is a qualitative evaluation of the performance of a new weighted statistic for a nonparametric test for stochastic dominance based on two samples, which was introduced in Ledwina and Wyłupek [*TEST* **21** (2012) 730–756]. For this purpose, we elaborate a useful variant of Kallenberg's notion of intermediate efficiency. This variant is general enough to be applicable to other nonparametric problems. We provide a formal definition of the proposed variant of intermediate efficiency, describe the technical tools used in its calculation, and provide proofs of related asymptotic results. Next, we apply this approach to calculating the intermediate efficiency of the new test with respect to the classical one-sided Kolmogorov–Smirnov test, which is a recognized standard for this problem. It turns out that for a very large class of convergent alternatives the new test is more efficient than the classical one. We also report the results of an extensive simulation study on the powers of the tests considered, which shows that the new variant of intermediate efficiency reflects the exact behavior of the power well.

Mathematics Subject Classification. 62G10, 62G20, 60E15.

Received February 9, 2018. Accepted October 26, 2018.

1. Introduction

In order to assess the performance of a test, a multitude of concepts of efficiency have been proposed. See Nikitin [42] and Serfling [46] for an overview of earlier definitions of asymptotic relative efficiencies. More recently, some efficiency measures have been defined in terms of probabilities of large and moderate deviations of type I and type II errors of tests; *cf.* Borovkov and Mogulskii [7], Ermakov [10, 11], and Kitamura [31] for some ideas, related history, and implementations. Obviously, each approach has its own rationale, some limitations, inherent complexity, and was developed with its own specific assumptions.

In this paper, we concentrate on the concept of asymptotic relative efficiency (ARE), and restrict our attention to Kallenberg's notion of intermediate efficiency. We develop a variant of this notion which is useful in

*Dedicated to Wilbert Kallenberg, with friendship and esteem.

Keywords and phrases: Anderson–Darling weight, asymptotic relative efficiency, goodness-of-fit, independence testing, Kallenberg efficiency, Kolmogorov–Smirnov test, local alternatives, moderate deviations, stochastic order, two-sample problem.

[1] Faculty of Pure and Applied Mathematics, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland.
[2] Institute of Mathematics, Polish Academy of Sciences, ul. Kopernika 18, 51-617 Wrocław, Poland.
[3] Faculty of Applied Mathematics, AGH University of Science and Technology, Al. Mickiewicza 30, 30-059 Cracov, Poland.

** Corresponding author: cmielbog@gmail.com

nonparametric problems and enables widely applicable, tractable, analytic comparisons between two tests. The results from earlier applications of Kallenberg's concept of efficiency are very encouraging. Kallenberg [28] gave several implementations of this notion to some popular tests for selected parametric and semiparametric models. Further developments include: adaptive test statistics [16, 22, 23], some classical goodness-of-fit statistics for both continuous and discrete data [23, 24], and [39], some classical statistics for testing a simple parametric hypothesis for a model of one-sample censorship [33], and testing for no-effect in certain regression models [24, 38].

In Section 3, we apply our variant of this approach to a recent solution of the one-sided nonparametric test for stochastic dominance introduced in Ledwina and Wyłupek [35]. Section 5.4.3 of that paper contains an extensive discussion motivating such an investigation. In fact, here we consider a generalization of that solution and denote it by $\mathcal{T}_N$; *cf.* (3.8). The test statistic $\mathcal{T}_N$ is asymptotically equivalent to the maximum of a weighted rank empirical process over a grid in (0,1), where the end points depend on the sample size and the classical Anderson–Darling weight is used. This last statistic is denoted by $\mathcal{W}_N$; see (3.9). We compare the new solution with the classical (unweighted) two-sample Kolmogorov–Smirnov test $\mathcal{V}_N$, which is commonly applied to this problem. In order to obtain a formula for the efficiency, several technical results have to be proved. Bounds for the asymptotic power of $\mathcal{T}_N$, under sequences of alternatives, and moderate deviations under the appropriate null distributions, are the main technical results of this paper. We prove that, for this one-sided nonparametric test, carefully matching the weight and the range of the maximum in $\mathcal{W}_N$ is highly profitable and results in a test which dominates the classical unweighted solution, regardless of how the alternative deviates from the null hypothesis. In particular, our Theorem 3.5 shows that the intermediate efficiency of the new test with respect to the Kolmogorov–Smirnov test is greater or equal than 1 for a very large class of sequences of alternatives. In simple terms, we compare two consistent tests which, in principle, can detect any fixed alternative. The only question is how many observations are needed for them to attain a given power. The intermediate efficiency of a weighted test relative to an unweighted one is the number which indicates by which factor one has to increase the sample size when using the unweighted solution to have approximately the same power as the weighted procedure has. Our simulations illustrate that the value of this efficiency measure appropriately explains the empirical behavior of these tests' power in this sense.

We close the introduction with a brief discussion of Kallenberg's notion of ARE and some related problems. The efficiency measure was introduced by matching the basic features of the approaches of Pitman and Bahadur to ARE. This involves alternatives converging to the null model (slower than in Pitman's approach) and significance levels tending to 0 (slower than in Bahadur's theory). Obviously, according to this approach, both the significance level and the alternative depend on the sample size N. The essence of this setting is that the significance level goes to 0 as N increases, while the asymptotic power, under the underlying sequence of local alternatives, should be non-degenerate. Hence, the sum of the type I and type II errors is in (0,1). Such requirements call for a careful and delicate balancing of the rates at which the significance levels tend to 0 and the local alternatives approach the null model. Although Kallenberg's concept of efficiency is slightly more complicated than the classical notions of Pitman and Bahadur, it is more widely applicable. For further discussion see Section 2.5.

The paper of Kallenberg [28] is mathematically elegant. Its main part concerns the one-sample case. In addition, it puts emphasis on having results which hold for all possible sequences of local alternatives. These rather stringent conditions can result in it being impossible to implement, even in relatively simple situations. Moreover, one important question regarding ensuring the non-degeneracy of the asymptotic power was skipped in all the examples given in that paper, while the k-sample case was not described in detail. Inglot [16] provided further analysis of Kallenberg's efficiency in the one-sample case, mainly in the context of studying adaptive Neyman tests. Our Remark 2.7 in Section 2 carefully discusses the differences between Kallenberg's original approach and our contribution.

In view of the above, in the present paper we propose a simple as possible variant of the notion of intermediate efficiency and define tools which help to calculate it. In contrast to the original concept, we do not require that results hold for all local alternative sequences. Our focus is on relaxing the requirements on the test statistics applied by as much as possible. Moreover, we embed the one- and two-sample cases into a joint scheme and study them simultaneously. Our contribution is self-contained: all of the necessary tools are carefully described, sufficient conditions for the non-degeneracy of the limiting power are given, and thorough proofs of all the related results are provided. The details are presented in Section 2, as well as Appendices A and B.

The organization of the paper is as follows: Section 2 describes our setup, the pathwise variant of intermediate efficiency, and discusses previous work in the light of our proposals. After this preparatory material, Section 3 presents two constructions of tests (a new and a classical one) for detecting stochastic ordering, introduces a natural class of sequences of alternatives for this problem, gives the theoretical results needed to compare such tests *via* intermediate ARE, and gives an explicit formula for the intermediate efficiency of the new solution with respect to the classical Kolmogorov–Smirnov test. We conclude by reporting some results of an extensive simulation study, which aims to show the usefulness of this variant of ARE when evaluating powers for finite samples. The proofs of all the technical results are given in Appendices A–G. Appendix B also provides some technical lemmas, which are useful in checking the assumptions of Theorem 2.6, the main theorem.

2. Intermediate efficiency, pathwise variant, basic facts and comments

2.1. Notation and definitions

Consider two independent samples $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ defined on the same measurable space and coming from the probability distributions P and Q, respectively. The situation $P = Q$ corresponds to the one-sample case.

Assume that $m = m(N)$, $n = n(N)$, with $N = m + n$. In the one-sample case, one can assume that the sample size is N, $m = m(N) = \lfloor N/2 \rfloor$ and $n = N - m$. Suppose that $\eta_N = m(N)/N$ satisfies $\eta_N \to \eta \in (0, 1)$, as $N \to \infty$. Denote the set of all product measures $P \times Q$ under consideration by $\mathbb{P}$. Let $\mathbb{P}_0$ be a subset of $\mathbb{P}$. We want to test whether

$$\mathbb{H}_0 : P \times Q \in \mathbb{P}_0 \quad \text{or} \quad \mathbb{H}_1 : P \times Q \in \mathbb{P}_1 = \mathbb{P} \setminus \mathbb{P}_0$$

is true.

Suppose that we have two upper-tailed tests defined by sequences of real valued test statistics $\mathcal{U}_N^{(\mathrm{I})}$ and $\mathcal{U}_N^{(\mathrm{II})}$ and critical values corresponding to a significance level $\alpha \in (0, 1)$ and sample size N, denoted by $u_{\alpha N}^{(\mathrm{I})}$ and $u_{\alpha N}^{(\mathrm{II})}$, respectively. To be specific, let

$$u_{\alpha N}^{(\mathrm{I})} = \inf\Big\{ w : \sup_{P \times Q \in \mathbb{P}_0} P^{m(N)} \times Q^{n(N)}\big(\mathcal{U}_N^{(\mathrm{I})} > w\big) \le \alpha \Big\},$$

where $P^{m(N)}$ and $Q^{n(N)}$ are $m(N)$ and $n(N)$ fold products of P and Q, respectively. We define $u_{\alpha N}^{(\mathrm{II})}$ in an analogous way. Throughout this article, $\{s_N\}$ denotes an infinite sequence of elements s_N, $N \ge 1$.

We wish to evaluate the efficiency of the test based on $\mathcal{U}^{(\mathrm{II})} = \{\mathcal{U}_N^{(\mathrm{II})}\}$ by comparing its sensitivity relative to the sensitivity of the test based on $\mathcal{U}^{(\mathrm{I})} = \{\mathcal{U}_N^{(\mathrm{I})}\}$. Hence, the test based on $\mathcal{U}^{(\mathrm{I})}$ is used as a benchmark.

We shall consider significance levels which tend to 0 as the sample size grows. Namely, the set $\mathbb{L}$ of sequences of all admissible levels in the intermediate setting is defined by

$$\mathbb{L} = \big\{ \{\alpha_N\} : \alpha_N \to 0, \;\; N^{-1} \log \alpha_N \to 0 \big\}. \tag{2.1}$$

Note that (2.1) excludes significance levels α_N which tend to 0 exponentially fast.

Given $P_0 \times Q_0 \in \mathbb{P}_0$, $P_1 \times Q_1 \in \mathbb{P}_1$, and $\vartheta \in (0, 1)$, define

$$P_\vartheta = (1 - \vartheta) P_0 + \vartheta P_1, \qquad Q_\vartheta = (1 - \vartheta) Q_0 + \vartheta Q_1, \tag{2.2}$$

and assume that

$$P_\vartheta \times Q_\vartheta \in \mathbb{P}_1 \quad \text{for every } \vartheta \in (0,1). \tag{2.3}$$

Hereafter, the measures $P_0 \times Q_0 \in \mathbb{P}_0$ and $P_1 \times Q_1 \in \mathbb{P}_1$ defining $P_\vartheta \times Q_\vartheta$ and satisfying (2.3) are fixed. Remark 2.2 gives illustrative constructions of P_ϑ and Q_ϑ in standard nonparametric testing problems.

Now, consider a particular sequence $\{\theta_N\}$, $\theta_N \in (0,1)$, where $\theta_N \to 0$, as $N \to \infty$, and the corresponding sequence $\{P_{\theta_N} \times Q_{\theta_N}\}$. Given N and the related θ_N, for an arbitrary natural number S representing the total size of an auxiliary sample and corresponding $m(S)$ and $n(S)$, define $\Pi^S_{\theta_N} = P^{m(S)}_{\theta_N} \times Q^{n(S)}_{\theta_N}$. Thus $\Pi^N_{\theta_N} = P^{m(N)}_{\theta_N} \times Q^{n(N)}_{\theta_N}$. Finally, set

$$\pi = \left\{\Pi^N_{\theta_N}\right\}.$$

Suppose that there exists $\{\alpha_N\} = \{\alpha_N(\pi)\} \in \mathbb{L}$ which satisfies

$$0 < \liminf_{N\to\infty} \Pi^N_{\theta_N}\left(\mathcal{U}^{(\mathrm{II})}_N > u^{(\mathrm{II})}_{\alpha_N N}\right) \le \limsup_{N\to\infty} \Pi^N_{\theta_N}\left(\mathcal{U}^{(\mathrm{II})}_N > u^{(\mathrm{II})}_{\alpha_N N}\right) < 1. \tag{2.4}$$

Let

$$\mathbb{L}^* = \mathbb{L}^*(\pi) = \Big\{\{\alpha_N\} \in \mathbb{L} : (2.4) \text{ holds}\Big\}. \tag{2.5}$$

In consequence, given $\{\theta_N\}$, for all $\{\alpha_N\} \in \mathbb{L}^*$, the corresponding test based on $\mathcal{U}^{(\mathrm{II})}_N$ has non-degenerate asymptotic power under $\{P_{\theta_N} \times Q_{\theta_N}\}$. In the sequel, we assume that $\mathbb{L}^*$ is nonempty.

The definition of the intermediate efficiency of $\mathcal{U}^{(\mathrm{II})}$ with respect to $\mathcal{U}^{(\mathrm{I})}$, which we give below, refers to this particular sequence of alternatives $\{P_{\theta_N} \times Q_{\theta_N}\}$, together with the corresponding π, and sequences of significance levels $\{\alpha_N\} \in \mathbb{L}^*$. For every $N \ge 1$, let

$$M_{\mathcal{U}^{(\mathrm{II})}\mathcal{U}^{(\mathrm{I})}}(N, \pi) = \inf\Big\{M \ge 1 : \Pi^N_{\theta_N}\left(\mathcal{U}^{(\mathrm{II})}_N > u^{(\mathrm{II})}_{\alpha_N N}\right) \le \Pi^{M+k}_{\theta_N}\left(\mathcal{U}^{(\mathrm{I})}_{M+k} > u^{(\mathrm{I})}_{\alpha_N M+k}\right) \ \text{for all } k \ge 0\Big\}. \tag{2.6}$$

Definition 2.1. If

$$e_{\mathcal{U}^{(\mathrm{II})}\mathcal{U}^{(\mathrm{I})}} = \lim_{N\to\infty} \frac{M_{\mathcal{U}^{(\mathrm{II})}\mathcal{U}^{(\mathrm{I})}}(N, \pi)}{N} \in [0, \infty] \tag{2.7}$$

exists and does not depend on the choice of $\{\alpha_N\} \in \mathbb{L}^*$, we say that the asymptotic intermediate efficiency of $\mathcal{U}^{(\mathrm{II})}$ with respect to $\mathcal{U}^{(\mathrm{I})}$, under the sequence of alternatives $\{P_{\theta_N} \times Q_{\theta_N}\}$, exists and equals $e_{\mathcal{U}^{(\mathrm{II})}\mathcal{U}^{(\mathrm{I})}}$.

Obviously, the asymptotic behavior of $M_{\mathcal{U}^{(\mathrm{II})}\mathcal{U}^{(\mathrm{I})}}(N, \pi)$, and hence $e_{\mathcal{U}^{(\mathrm{II})}\mathcal{U}^{(\mathrm{I})}}$, depends on η. Typically, the value of $\eta \in (0,1)$ is fixed and therefore, to simplify the notation, this parameter is omitted in (2.7). However, in Section 4 we analyze numerically the influence of η on the efficiency of tests for the two-sample case. Therefore, in Theorem 3.5, which presents an analytic formula for this measure of efficiency, and in Section 4 we clearly indicate the dependence of efficiency on η.

In contrast to the original definition of Kallenberg [28] and its extension by Inglot [16], where a counterpart of (2.7) was required for some families of sequences of alternatives, the above definition is restricted to a particular sequence. Hence, it can be considered to be a kind of pathwise variant of the previous approaches. The path π is uniquely determined *via* $P_0 \times Q_0$, $P_1 \times Q_1$, $\{\theta_N\}$, $\{m(N)\}$ and $\{n(N)\}$. Such a pathwise approach extends the range of possible applications of the notion and allows us to avoid many non-trivial technicalities. In particular, we avoid the introduction of so-called renumerable families, which are key objects in Inglot [16]. Note also that, as a rule, (2.7) holds for many sequences simultaneously, but the above definition treats each of them separately. For an illustration of this, see Theorem 3.5 and Remark 3.6 following it.

Remark 2.2. This remark gives illustrative constructions of P_ϑ and Q_ϑ in standard nonparametric problems. As typical in such problems, we shall start with underlying cumulative distribution functions of observed random variables and explain details in such terms. Latter on related P_ϑ's and Q_ϑ's shall be introduced. Typical property of standard nonparametric problems is also their invariance to monotonic transformations. Such transformations are very useful to see an essence of the problem and we shall apply some transformations to provide statistical interpretation of (2.2).

Consider first one-sample case. Let $Z_1, \ldots, Z_N$ be independent real valued random variables with continuous distribution function F. Suppose that we like to verify simple goodness-of-fit hypothesis $F = F_0$ against a class of mixtures $(1-\vartheta)F_0 + \vartheta F_1,\ \vartheta \in (0,1)$. Detecting mixtures is hot topic recent years. Moreover, always such structure can be interpreted as a way of approaching F_0 from 'the direction' F_1. Then, P_1 and P_0 are simply related probability distributions and $P_\vartheta = (1-\vartheta)P_0 + \vartheta P_1$. Alternatively, one can consider the transformed sample $F_0(Z_1), \ldots, F_0(Z_N)$. Then, under the null hypothesis, the random variable $F_0(Z_1)$ has uniform probability distribution $\bar{P}_0$, say, while P_ϑ transforms to $\bar{P}_\vartheta = (1-\vartheta)\bar{P}_0 + \vartheta\bar{P}_1$, where $\bar{P}_1$ is a probability distribution pertaining to $F_1 \circ F_0^{-1}$. The last mentioned function is called the relative distribution of F_1 to F_0 and plays important role in many areas of applied statistics.

In the case of independence testing consider random vectors $Z_1, \ldots, Z_N$ with continuous marginal distribution functions. Then, after imposing the marginal distribution functions on the components of Z_i's, without loss of generality, one can assume that the marginals are uniform on [0,1] and restrict attention to the unique copula associated with joint cumulative distribution of Z_1, say; see Fermanian *et al.* [13] for more details. It is widely accepted opinion that the dependence structure among the components of the observed random vector is best characterized by its unique copula. After such reparametrization, a copula is a parameter under test while marginal distributions are nuisance parameters. Then (2.2) can be interpreted as a mixture of the independence copula and some alternative copula. To be specific, let C_0 stands for the independence copula, *i.e.* uniform distribution function on the unit square, while C_1 is a copula representing an alternative distribution. Then P_ϑ is pertaining to the mixture $(1-\vartheta)C_0 + \vartheta C_1$.

Two-sample problem is the most difficult for an insightful reparametrization. A breakthrough in this question was the paper by Behnen [4], which we shall use below. See also Behnen and Neuhaus [5], and Neuhaus [41] for related information. Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be two independent samples with respective parent continuous cumulative distribution functions F and G. As before, we set $N = m + n$ and assume that $\eta = \lim(m/N)$ exists and $\eta \in (0,1)$. Moreover, let $Z_1, \ldots, Z_N$ represent the pooled sample. For simplicity, we focus here on testing $\mathbb{H}_0 : F = G$ against $\mathbb{H}_1 : F \neq G$. For related comments on testing the stochastic dominance see Section 3. To define (2.2) for $\mathbb{H}_0$ against $\mathbb{H}_1$, in contrast to the two above examples, in this setting we start from an alternative distribution. So, take arbitrary two continuous distribution functions F_1, G_1, say, where $F_1 \neq G_1$. Now, define F_0, say, by setting $F_0(z) = J_1(z) = \eta F_1(z) + (1-\eta)G_1(z),\ z \in \mathbb{R}$. The probability measure corresponding to this F_0 is denoted by P_0. With the above choice of F_0, for a given real number $\vartheta \in (0,1)$, we introduce a contamination model $F_\vartheta = (1-\vartheta)F_0 + \vartheta F_1, G_\vartheta = (1-\vartheta)F_0 + \vartheta G_1$. We denote by P_ϑ and Q_ϑ the probability measures corresponding to F_ϑ and G_ϑ.

To have some more clear link to the previous cases, observe that the definition of J_1 implies that $F_1 = J_1 + (1-\eta)(F_1 - G_1),\ G_1 = J_1 - \eta(F_1 - G_1)$. Hence, transformed by $F_0 = J_1$ observations $Z_1, \ldots, Z_N$, obey, under $\mathbb{H}_0$, the uniform (0,1) distribution, while under $\mathbb{H}_1$ their distribution functions in the first and the second sample, respectively, take the form

$$\Phi_1(t) = t + (1-\eta)(F_1 - G_1) \circ J_1^{-1}(t), \quad \Phi_2(t) = t - \eta(F_1 - G_1) \circ J_1^{-1}(t), \quad \text{where} \quad t \in (0,1).$$

In this parametrization $(F_1 - G_1) \circ J_1^{-1}$ is the parameter under test while J_1 is a nuisance parameter. Hence, in the transformed samples we have $\bar{P}_\vartheta = (1-\vartheta)\bar{P}_0 + \vartheta\bar{P}_1$ and $\bar{Q}_\vartheta = (1-\vartheta)\bar{P}_0 + \vartheta\bar{Q}_1$, where the probability measures $\bar{P}_0,\ \bar{P}_1$ and $\bar{Q}_1$ are pertaining to the uniform, Φ_1 and Φ_2 distribution functions, accordingly. Here

$F_1 \circ J_1^{-1}$ and $G_1 \circ J_1^{-1}$ are Parzen's relative distributions of F_1 and G_1 to J_1. See Handcock and Morris [15] for details on relative distributions. A great advantage of the above reparametrization is that the function $\bar{A}(t) = (F_1 - G_1) \circ J_1^{-1}(t)$ is absolutely continuous and its derivative $\bar{a}(t)$, say, is bounded. This greatly simplifies asymptotic derivations under related sequences of alternatives, defined in Section 3.1.

The k-sample problem can be treated completely analogously. For further details see also Behnen and Neuhaus [6].

Remark 2.3. We can succinctly rephrase the interpretation of positive and finite $e_{\mathcal{U}^{(\mathrm{II})}\mathcal{U}^{(\mathrm{I})}}$ as follows: the test corresponding to $\mathcal{U}^{(\mathrm{I})}$ and related to the sample sizes $\lfloor me_{\mathcal{U}^{(\mathrm{II})}\mathcal{U}^{(\mathrm{I})}} \rfloor$ and $\lfloor ne_{\mathcal{U}^{(\mathrm{II})}\mathcal{U}^{(\mathrm{I})}} \rfloor$ has approximately the same power, under $\{P_{\theta_N} \times Q_{\theta_N}\}$, as the power of the test corresponding to $\mathcal{U}^{(\mathrm{II})}$ and pertaining to the sample sizes m and n.

Another useful interpretation of the intermediate efficiency is the value of the shift of a non-parametric alternative necessary for the two tests under consideration to have the same local power. See Inglot and Ledwina [23] for some simple illustration and Inglot [18] for further development.

Remark 2.4. To prove that the limit in (2.7) exists and to obtain an explicit formula for it, we need to introduce some regularity assumptions for both test statistics. Sequences of statistics satisfying assumptions of this kind are called Kallenberg sequences by Koning [33]. Similarly to Inglot [16] and in contrast to Kallenberg [28], we impose stronger requirements on $\mathcal{U}^{(\mathrm{I})}$ than on $\mathcal{U}^{(\mathrm{II})}$. On one hand, the benchmark, $\mathcal{U}^{(\mathrm{I})}$, can be always chosen in a convenient way. On the other hand, any relaxation of the requirements on $\mathcal{U}^{(\mathrm{II})}$ extends the scope of possible applications of this approach. Similar to the Bahadur efficiency, the intermediate efficiency of $\mathcal{U}^{(\mathrm{II})}$ with respect to $\mathcal{U}^{(\mathrm{I})}$ is calculated as the ratio between two slopes. These slopes are determined by an index for moderate deviations under the null hypothesis and a scaling factor which results from a kind of weak law of large numbers (WLLN) under the sequence of alternatives. It is worth emphasizing that we only assume a knowledge of moderate deviations of $\mathcal{U}^{(\mathrm{II})}$ in some restricted range. In contrast, full range of moderate deviations of $\mathcal{U}^{(\mathrm{I})}$ is required.

2.2. Regularity assumptions on $\mathcal{U}^{(\mathrm{I})}$

(I.1) There exists a positive number $c_{\mathcal{U}^{(\mathrm{I})}}$ such that for every positive sequence $\{w_N\}$ satisfying $w_N \to 0$ and $Nw_N^2 \to \infty$, the following holds:

$$- \lim_{N\to\infty} \frac{1}{Nw_N^2} \log \sup_{P\times Q \in \mathbb{P}_0} P^{m(N)} \times Q^{n(N)} \big(\mathcal{U}_N^{(\mathrm{I})} > w_N \sqrt{N}\big) = c_{\mathcal{U}^{(\mathrm{I})}}.$$

(I.2) There exists a function $b_{\mathcal{U}^{(\mathrm{I})}}(P_\vartheta \times Q_\vartheta)$, $\vartheta \in (0,1)$, and a number $\rho \in [1,2]$, such that for every sequence $\{\vartheta_N\}$ of positive numbers satisfying $\vartheta_N \to 0$ and $N\vartheta_N^\rho \to \infty$ it holds that

$$\lim_{N\to\infty} \Pi_{\vartheta_N}^N \left(\left| \frac{\mathcal{U}_N^{(\mathrm{I})}}{\sqrt{mn/N}\, b_{\mathcal{U}^{(\mathrm{I})}}(P_{\vartheta_N} \times Q_{\vartheta_N})} - 1 \right| \geq \epsilon \right) = 0 \quad \text{for every} \quad \epsilon > 0, \tag{2.8}$$

where $\Pi_{\vartheta_N}^N = P_{\vartheta_N}^{m(N)} \times Q_{\vartheta_N}^{n(N)}$.

We call $c_{\mathcal{U}^{(\mathrm{I})}}$ the index of moderate deviations of $\mathcal{U}^{(\mathrm{I})}$, while the quantity $c_{\mathcal{U}^{(\mathrm{I})}}\big[b_{\mathcal{U}^{(\mathrm{I})}}(\Pi_{\vartheta_N}^N)\big]^2$, where $b_{\mathcal{U}^{(\mathrm{I})}}(\Pi_{\vartheta_N}^N) = \sqrt{\frac{mn}{N}} b_{\mathcal{U}^{(\mathrm{I})}}(P_{\vartheta_N} \times Q_{\vartheta_N})$, shall be called the intermediate slope of $\mathcal{U}^{(\mathrm{I})}$ under $\Pi_{\vartheta_N}^N$.

2.3. Regularity assumptions on $\mathcal{U}^{(\mathrm{II})}$

(II.1) There exist sequences $\{\gamma_N\}$ and $\{\lambda_N\}$, such that $1 \leq \gamma_N < \lambda_N \leq N$, $\gamma_N/\lambda_N \to 0$, and for every positive sequence $\{w_N\}$, where $Nw_N^2/\lambda_N \to 0$ and $Nw_N^2/\gamma_N \to \infty$, the following holds:

$$- \lim_{N\to\infty} \frac{1}{Nw_N^2} \log \sup_{P\times Q \in \mathbb{P}_0} P^{m(N)} \times Q^{n(N)} \big(\mathcal{U}_N^{(\mathrm{II})} > w_N \sqrt{N}\big) = c_{\mathcal{U}^{(\mathrm{II})}} \tag{2.9}$$

for a positive number $c_{\mathcal{U}^{(\mathrm{II})}}$.

(II.2) For a particular sequence $\{\theta_N\}$, defining the sequence $\{P_{\theta_N} \times Q_{\theta_N}\}$ of alternatives under consideration, and such that

$$\lim_{N\to\infty} \Pi^N_{\theta_N}\Big(\Big|\frac{\mathcal{U}_N^{(\mathrm{II})}}{b_{\mathcal{U}^{(\mathrm{II})}}(\Pi^N_{\theta_N})} - 1\Big| \geq \epsilon\Big) = 0 \quad \text{for every } \ \epsilon > 0. \tag{2.10}$$

As above, the quantity $c_{\mathcal{U}^{(\mathrm{II})}}\big[b_{\mathcal{U}^{(\mathrm{II})}}(\Pi^N_{\theta_N})\big]^2$ shall be called the intermediate slope of $\mathcal{U}^{(\mathrm{II})}$ under $\Pi^N_{\theta_N}$.

Remark 2.5. In (I.2) and (II.2), we imposed the assumption $N\theta_N^{\rho} \to \infty$, for some $\rho \in [1,2]$. In several previously considered cases, it was simply assumed that $\rho = 2$. As seen from the proof of Theorem 2.6, given in Appendix A, this assumption is closely related to the behavior of the probability of a type one error in the test under consideration, which in turn depends explicitly on the intermediate slope. Obviously, the most demanding conditions are provided by the Neyman–Pearson test. The intermediate slope of the Neyman–Pearson test for two-sample problems was studied in Ducharme and Ledwina (*cf.* [9] Thms. 3.3 and 3.4) In this case, $\rho = 2$ is appropriate. In other tests, the situation might be different.

2.4. Computation of the intermediate efficiency

We start with a counterpart to Lemmas 2.1 and 5.1 in Kallenberg [28]. This result gives conditions under which the asymptotic ratio of the slopes of $\mathcal{U}^{(\mathrm{I})}$ and $\mathcal{U}^{(\mathrm{II})}$ coincides with the asymptotic ratio of the sample sizes in (2.7). Next, we compare our result to the lemmas mentioned above and comment on how the assumptions of our result can be checked.

Theorem 2.6. *Assume that for a given $\mathcal{U}^{(\mathrm{I})}$ the conditions (I.1) and (I.2) are true for some $\rho \in [1,2]$. Suppose that, for a particular sequence $\{\theta_N\}$, $\mathcal{U}^{(\mathrm{II})}$ satisfies (II.2). Moreover, (II.1) holds for some given $\{\gamma_N\}$ and $\{\lambda_N\}$. Suppose that there exists $\{\alpha_N\}$ from $\mathbb{L}^*$ such that*

$$\lim_{N\to\infty} \frac{\log \alpha_N}{\lambda_N} = 0 \quad \textit{and} \quad \lim_{N\to\infty} \frac{\log \alpha_N}{\gamma_N} = -\infty. \tag{2.11}$$

Then it holds that
(i) $-\log \alpha_N = c_{\mathcal{U}^{(\mathrm{II})}}[b_{\mathcal{U}^{(\mathrm{II})}}(\Pi^N_{\theta_N})]^2[1+o(1)]$,
(ii) for all $\{\alpha'_N\} \in \mathbb{L}^*$ *we have* $\lim_{N\to\infty} \log \alpha'_N / \log \alpha_N = 1$ *what means that every sequence form* $\mathbb{L}^*$ *fulfills (2.11) as well.*

Finally assume that the following limit exists

$$\lim_{N\to\infty} \frac{c_{\mathcal{U}^{(\mathrm{II})}}[b_{\mathcal{U}^{(\mathrm{II})}}(\Pi^N_{\theta_N})]^2}{c_{\mathcal{U}^{(\mathrm{I})}}(mn/N)[b_{\mathcal{U}^{(\mathrm{I})}}(P_{\theta_N} \times Q_{\theta_N})]^2} = \mathbf{e} \in [0,\infty]. \tag{2.12}$$

Then the intermediate efficiency (2.7) of $\mathcal{U}^{(\mathrm{II})}$ with respect to $\mathcal{U}^{(\mathrm{I})}$, under a particular sequence of alternatives $\{P_{\theta_N} \times Q_{\theta_N}\}$, exists and $e_{\mathcal{U}^{(\mathrm{II})}\mathcal{U}^{(\mathrm{I})}} = \mathbf{e}$.

Let us start by discussing the differences between our requirements and those imposed in Lemmas 2.1 and 5.1 of Kallenberg [28].

Remark 2.7. The first essential difference between our paper and Kallenberg's one consists in the fact that Kallenberg [28] requires that for both statistics $\mathcal{U}_N^{(\mathrm{I})}$ and $\mathcal{U}_N^{(\mathrm{II})}$ the WLLN under the alternatives holds for all allowable sequences $\{\theta_N\}$ and all corresponding distributions, provided that his condition (1.3) holds. This can be an inappropriate formulation and his paper provides examples where some further restrictions are imposed, and, in fact, some paths introduced.

The second essential difference between both papers lies in the fact that Kallenberg [28] assumes the same type of moderate deviations for both $\mathcal{U}_N^{(\mathrm{I})}$ and $\mathcal{U}_N^{(\mathrm{II})}$, and does not impose any restriction on the rate of convergence of $\{w_N\}$ (in our notation) from above. In contrast, we introduce such a requirement in the case of $\mathcal{U}_N^{(\mathrm{II})}$, and this is expressed by the assumption that $Nw_N^2/\gamma_N \to \infty$. We impose such an assumption because it naturally arises when studying the statistic $\mathcal{T}_N$; *cf.* Theorem 3.2. Such an assumption is simply indispensable in many cases, as, for example, for some weighted goodness of fit statistics where moderate deviations exist and are only non-zero for a very restricted type of sequences of $\{w_N\}$'s. To balance this useful restriction on moderate deviations for $\mathcal{U}_N^{(\mathrm{II})}$, we assume that $\mathcal{U}_N^{(\mathrm{I})}$ has non-zero moderate deviations for the whole range of the w_N's. This is not a restrictive assumption, because any convenient benchmark $\mathcal{U}_N^{(\mathrm{I})}$ can be chosen.

The third difference is that we do not require that $b_{\mathcal{U}^{(\mathrm{II})}}(\Pi_{\theta_N}^N)$ has a special structure. This allows us to compare statistics with a different rate of convergence than the one corresponding to $\mathcal{U}_N^{(\mathrm{I})}$.

Additionally, our paper clarifies the intermediate approach for the two-sample case. Note that, putting aside the limitations discussed above, the formulation of (iii) in Kallenberg's Lemma 5.1 contains some further non-explicit restrictions.

Finally, observe that the essential assumption regarding the non-degeneracy of the asymptotic power of $\mathcal{U}_N^{(\mathrm{II})}$, in our notation, is missing in the formulations of Kallenberg's lemmas, though it is clearly stated on page 171 of Kallenberg [28]. This may be misleading, as it is needed in the proof. Verification of this assumption is a non-trivial problem, which is ignored in the analysis of the examples in Kallenberg [28]. In our Theorem 2.6, this assumption is implicit in the condition $\{\alpha_N\} \in \mathbb{L}^*$, and later we provide some convenient tools for checking this assumption; *cf.* Remark 2.8.

Now, we give some brief comments on verifying the assumptions of our Theorem 2.6.

Remark 2.8. The regularity assumptions (I.1) and (II.1) hold for many classical statistics. In the present paper, the two-sample Kolmogorov–Smirnov statistic serves as a good example. Often, the strong approximations method proves to be very useful in obtaining such results.

For some statistics, conditions like (I.2) and (II.2), together with the form of α_N and non-degeneracy of asymptotic powers, have already been justified on the basis of the limiting distribution of the underlying test statistic under given sequences of alternatives; *cf.* Inglot and Ledwina [22, 23] for some examples. However, such a limiting distribution is often hard to derive. Therefore, following the idea applied for the first time in Inglot and Ledwina [24], we propose to use some asymptotic bounds for the distributions of test statistics under the considered sequence of alternatives. These bounds have the condition of non-degenerate asymptotic powers explicitly built into them, which is very useful in checking (I.2) and (II.2) in some relatively complex cases. For details, see Lemmas B.1 and B.2 in Appendix B, where we also give an example of a sequence $\{\alpha_N\} \in \mathbb{L}^*$ and a range of non-degenerate asymptotic powers for which Theorem 2.6 applies. Note also that, by (ii) of Theorem 2.6, if we know at least one example of $\{\alpha_N\} \in \mathbb{L}^*$ then, in principle, we know all other sequences from $\mathbb{L}^*$. See also Sections 3.3–3.5 for an illustration of how works the approach described above.

We close this section with a discussion on other selected notions of efficiency and their relations to Kallenberg's efficiency. This section has been added to the revised version at a request of one of referees.

2.5. Kallenberg's efficiency as a compromise implementation of Pitman's and Bahadur's ideas

Let us start with an informal introduction of the Pitman and the Bahadur efficiencies. We pattern notation after Kallenberg [29]. Both efficiencies can be described in terms of sample sizes $N(\alpha, \beta, \upsilon)$, say, needed to attain with given test on a level α a prescribed power β; $\beta \in (0, 1)$ at an alternative υ. We do not specify a character of υ at the moment. For two tests, say 1 and 2, with corresponding numbers N_1 and N_2, respectively, the ratio N_2/N_1 is called the relative efficiency of test 1 with respect to test 2. It is rather clear that it is extremely difficult to compute $N_i(\alpha, \beta, \upsilon),\ i = 1, 2$, even for very simple test statistics. Therefore, some asymptotic approaches, as the sample size N tends to infinity, have been proposed. In particular, Pitman's approach assumes that $\upsilon = \upsilon_N$ and υ_N is getting closer to a null hypothesis as N grows. In contrast, Bahadur's idea was to set $\alpha = \alpha_N$ and send it to 0 as N increases. In both cases the power β was fixed. Both approaches are attractive. However, their implementations require different assumptions and have strong limitations. It is beyond the scope of this section to present some general results. Therefore, for an illustration, we shall mainly concentrate on pertaining problems in the case of weighted and unweighted Kolmogorov–Smirnov type tests for two-sample problem.

The Pitman efficiency was introduced in 1948 in unpublished notes; Pitman [43, 44]. It has been originally designed for parametric testing problems. See Nikitin [42] and Zacks [52] for insightful introduction and related literature. In general, this approach is linked to the central limit theorem and works nicely for asymptotically normal statistics. Otherwise, may not exist or may depend on the significance level α and given power β. A good illustration of these drawbacks are efforts to evaluate Pitman's efficiency of classical Kolmogorov–Smirnov test for two-sample problem with respect to some linear statistics; see Kalish and Mikulski [27], and Yu [51]. These papers deal with the case of location model indexed by $\upsilon \in R$. In particular, Kalish and Mikulski [27] have defined generalized Pitman efficiency while Yu [51] provided complicated bounds for the efficiency. Neuhaus [40] treated much more general nonparametric two-sample setting and derived asymptotic results allowing for some extension of the above mentioned results of Yu [51]. Neuhaus [40] defined also a sample version of Pitman's efficiency for nonparametric two-sample problem, *i.e.* for υ being infinite dimensional. To be specific, in terms of the present paper, his approach results in sequences of alternatives $\{P_{\theta_N} \times Q_{\theta_N}\}$, where $\theta_N = O(N^{-1/2})$ while P_{θ_N} and Q_{θ_N} are as in our Remark 2.2. The precise definition of the above mentioned $N(\alpha, \beta, \upsilon)$ in Neuhaus [40] is slightly less demanding than ours (2.6) while his definition (3.6') of the Pitman efficiency of Kolmogorov–Smirnov test with respect to linear rank tests is weaker than our general definition (2.7) in particular in that he uses *lim inf* (as in the generalized Pitman efficiency introduced by Kalish and Mikulski [27]) instead of our *lim*. No explicit analytical formula for this (generalized) Pitman efficiency (3.6') is provided in Neuhaus [40], though it is clear that the resulting quantity shall depend on α. Neuhaus's extension of the Pitman efficiency works nicely for comparison of two linear rank statistics and provides very welcome finite sample interpretation of earlier ARE notion based on asymptotic shifts; *cf.* his Example 3.2.

For parametric models indexed by a real parameter υ and simple null hypothesis $\upsilon = \upsilon_0$, say, Wieand [50] suggested a way of escaping from the problems appearing in Yu [51] and related papers. Namely, he introduced succeeding modification of the Pitman efficiency notion and proposed to calculate it under $\alpha \to 0$. The resulting quantity has been called limiting Pitman efficiency. For an application of this idea to comparisons of two-sided unweighted Kolmogorov–Smirnov two-sample test to some linear statistics, under location model, see Sinha and Wieand [48]. Ledwina [34] extended Wieand's approach to cover the case of nonparametric independence testing. Similar extension for other nonparametric problems is possible. Some other generalization was proposed in Kallenberg and Koning [30]. However, our main objection concerning this notion is its complex form, far from the basic idea of the Pitman efficiency. It seems that, Wieand's paper can be consider rather as a forerunner of more simple and intuitive intermediate efficiency notion than a generalization of Pitman's efficiency.

The exact Bahadur efficiency was proposed in Bahadur [1]. It concerns the situation when the alternative υ is fixed while the size of a test is decreasing exponentially fast as the number of observations N tends to infinity. An important advantage of Bahadur's efficiency consists in that it can be applied to statistics with non-normal asymptotic distributions such as unweighted Kolmogorov–Smirnov two-sample statistic. On the other hand, for calculating this efficiency it is necessary to have non-degenerate large deviation asymptotics of test statistic under the null hypothesis. This problem is always non-trivial but often also critical. For many statistics used at present, including several weighted ones, the large deviations degenerate, *i.e.* the index of large deviations is 0. For an illustration see Groeneboom and Shorack [14], Behnen and Neuhaus [5], and Inglot [17]. See also Section 4.3 of Jager and Wellner [26] as well as Jager and Wellner [25] for description and discussion of further problems with and application of Bahadur's approach to some weighted statistics. Hence, in particular, many weighted variants of Kolmogorov–Smirnov test are excluded from a consideration. For a derivation of Bahadur efficiency of unweighted two-sample Kolmogorov–Smirnov test with respect to some linear rank statistics see Nikitin [42].

This brief discussion shows that it is very welcome to have an ARE notion which shares advantages of the two above mentioned approaches and is feasible in the situations excluded by their limitations. Kallenberg's efficiency provides solution of such type. It assumes that the type I error tends to 0 (slower than exponentially), considers alternatives approaching the null model (slower than in the parametric rate $O(N^{-1/2})$) and requires only non-degenerate moderate deviations for the null distributions. So, it shares the essence of both celebrated approaches but allows for slower rates than in the source forms. Moreover, similarly as in the Bahadur approach, asymptotic normality of test statistics is not crucial. In comparison with Bahadur's ARE, the concept of intermediate efficiency requires similar, but less demanding conditions, to be applicable. Ermakov [11], page 622, admits: "Kallenberg's efficiency represents the analogue of the Bahadur efficiency in a moderate deviation zone". The paper of Kallenberg [28] discusses some relations between the above mentioned notions in the case of some parametric and semiparametric models. See also Inglot and Ledwina [20, 21] for some technical results helping to understand some equivalences between Bahadur's and Kallenberg's efficiency measures.

The above encourages us to paraphrase the opinion of Nikitin [42], page 11, on the Bahadur efficiency to the following one: a definite merit of the Kallenberg efficiency lies in the fact that it permits one to distinguish tests in the cases when other types of AREs are useless.

3. The intermediate efficiency of some Kolmogorov–Smirnov-type tests for stochastic ordering

In this section, step by step we shall develop the tools necessary to apply the results of Section 2 to some selected statistics for testing for the existence of stochastic dominance. In successive subsections we present this problem and the test statistics under consideration, introduce appropriate sequences of alternatives, collect some theoretical results on moderate deviations under the null hypothesis and asymptotic behavior under the sequences of alternatives. We conclude with Theorem 3.5, which states the existence and the form of the intermediate efficiency, and Corollary 3.8, which exemplifies an implementation of the general result.

3.1. The testing problem and local sequences of nonparametric alternatives

We consider two independent samples $X_1, \dots, X_m$ and $Y_1, \dots, Y_n$ which correspond to the continuous distribution functions F and G, respectively. As in Section 2.1, we assume that $m = m(N)$, $n = n(N)$, $m + n = N$, and both sample sizes tend to infinity as N increases. Moreover, we set $\eta_N = m(N)/N$ and suppose that

$$\eta = \lim_{N\to\infty} \eta_N \quad \text{exists and} \quad \eta \in (0, 1). \tag{3.1}$$

Throughout Section 3, all limits are taken under the assumption that m and n grow in such a way that (3.1) holds.

The null hypothesis, $\mathbb{H}_0$, asserts that the X's are stochastically smaller than the Y's, *i.e.*

$$\mathbb{H}_0 : F(z) \geq G(z) \quad \text{for each} \quad z \in \mathbb{R},$$

while the alternative, $\mathbb{H}_1$, is unrestricted and is of the form

$$\mathbb{H}_1 : F(z) < G(z) \quad \text{for some} \quad z \in \mathbb{R}.$$

Note that both the classical test for $\mathbb{H}_0$ and the new one, which we shall consider in the following subsections, are distribution free for any continuous $F = G$. Moreover, the following property holds:

$$Pr(\mathcal{S}_N > w|F \geq G) \leq Pr(\mathcal{S}_N > w|F = G), \quad \text{for all } \; w \in \mathbb{R},$$

where $\mathcal{S}_N$ is any of the considered test statistics. For details, see Ledwina and Wyłupek [35].

Here and throughout of the rest of the paper Pr denotes a probability on some probability space, perhaps different in each case.

The above suggests to restrict attention to a 'least favorable case' in $\mathbb{H}_0$, *i.e.* $F = G$, in construction of paths. Moreover, technical advantages also support such choice; see below. Therefore, to define a family of nonparametric paths, we proceed as described in Remark 2.2. See also Remark 2.4 in Neuhaus [40], Neuhaus [41] and Ducharme and Ledwina [9] for applications of such approach. Namely, we take an arbitrary alternative determined by the distribution functions F_1, G_1 and select $F = G = F_0 = J_1 = \eta F_1 + (1-\eta)G_1$. The probability measure corresponding to this F_0 is denoted by P_0. P_0^N denotes its N-fold product.

With the above choice of F_0, for a given sequence of real numbers $\vartheta_N \in (0, 1)$ such that $\vartheta_N \to 0$ as $N \to \infty$, we introduce a contamination model (F_{1N}, G_{1N}) based on

$$F_{1N} = (1 - \vartheta_N)F_0 + \vartheta_N F_1, \;\; G_{1N} = (1 - \vartheta_N)F_0 + \vartheta_N G_1. \tag{3.2}$$

We denote by P_{ϑ_N} and Q_{ϑ_N} the probability measures corresponding to F_{1N} and G_{1N}, defined in (3.2). The formula (3.2) defines our local sequence of nonparametric alternatives. Note that, due to the choice $F = G = F_0$ the formula (3.2) ensures (2.3). For other choices of the representative pair of null distributions such implication may not be true.

3.2. One-sided two-sample test statistics

The empirical distribution functions for the two samples are $\hat{F}_m(z) = m^{-1}\sum_{i=1}^m \mathbf{1}(X_i \leq z)$ and $\hat{G}_n(z) = n^{-1}\sum_{i=1}^n \mathbf{1}(Y_i \leq z)$, respectively, where $\mathbf{1}(\mathbb{E})$ is the indicator function of the set $\mathbb{E}$. Additionally, let $\hat{J}_N(z)$ be the empirical distribution function for the pooled sample, *i.e.*

$$\hat{J}_N(z) = \eta_N \hat{F}_m(z) + (1 - \eta_N)\hat{G}_n(z), \quad z \in \mathbb{R}, \tag{3.3}$$

and denote the left continuous inverse of $\hat{J}_N(z)$ by $\hat{J}_N^{-1}(t)$. To be specific, $\hat{J}_N^{-1}(t) = \inf\{z : \hat{J}_N(z) \geq t\}$ for $t \in (0, 1)$. The one-sided Kolmogorov–Smirnov test rejects $\mathbb{H}_0$ when

$$\mathcal{V}_N = \sqrt{\frac{mn}{N}} \sup_{z\in\mathbb{R}} \Big\{\hat{G}_n(z) - \hat{F}_m(z)\Big\} = \sqrt{\frac{mn}{N}} \max_{1\leq j\leq N} \Big\{\hat{G}_n - \hat{F}_m\Big\} \circ \hat{J}_N^{-1}\Big(\frac{j}{N}\Big) \tag{3.4}$$

exceeds the appropriate critical value. Note that $\mathcal{V}_N$ is a rank statistic.

For testing $\mathbb{H}_0$ against $\mathbb{H}_1$, Ledwina and Wyłupek [35] introduced, among other statistics, a test statistic based on the minimum of an appropriate set of linear rank statistics. Since $\mathbb{H}_0$ is one-sided, linear rank statistics with non-increasing score generating functions are appropriate; *cf.* Behnen [3]. In Ledwina and Wyłupek [35] step functions, related to projections of Haar functions, were used to define a set of useful rank statistics. Here, we slightly generalize this construction by allowing a more flexible set of step functions.

To be specific, let R_i, $i = 1, \ldots, m$, denote the rank of X_i in the pooled sample $X_1, \ldots, X_m, Y_1, \ldots, Y_n$. Analogously, R_i, $i = m+1, \ldots, N$, stands for the rank of Y_i in the pooled sample. Let $\Delta(N)$ be a non-decreasing sequence of natural numbers, such that $1 < \Delta(N) \le N$. Given $j = 1, \dots, \Delta(N)$, set

$$\ell_j(t) = \ell_{jN}(t) = -\sqrt{\frac{1-\pi_{jN}}{\pi_{jN}}}\,\mathbf{1}(0 \le t < \pi_{jN}) + \sqrt{\frac{\pi_{jN}}{1-\pi_{jN}}}\,\mathbf{1}(\pi_{jN} \le t \le 1), \tag{3.5}$$

where $0 < \pi_{1N} < \pi_{2N} < \cdots < \pi_{\Delta(N)N} < 1$.

Consider the corresponding linear rank statistics given by

$$\mathcal{L}_j = \mathcal{L}_{jN} = \sum_{i=1}^{N} c_{Ni}\,\ell_j\Big(\frac{R_i - 0.5}{N}\Big), \quad j = 1, \dots, \Delta(N), \tag{3.6}$$

where

$$c_{Ni} = \sqrt{\frac{mn}{N}} \begin{cases} -m^{-1} & \text{if} \quad 1 \le i \le m, \\ \;\; n^{-1} & \text{if} \quad m < i \le N. \end{cases} \tag{3.7}$$

From the above, after elementary calculations, we obtain

$$\mathcal{L}_j = \sqrt{\frac{mn}{N}}\frac{1}{\sqrt{\pi_{jN}(1-\pi_{jN})}} \times \left[\frac{1}{m}\sum_{i=1}^{m}\mathbf{1}_{[0,\pi_{jN})}\left(\frac{R_i-0.5}{N}\right) - \frac{1}{n}\sum_{i=m+1}^{N}\mathbf{1}_{[0,\pi_{jN})}\left(\frac{R_i-0.5}{N}\right)\right]$$

$$= \sqrt{\frac{mn}{N}}\frac{1}{\sqrt{\pi_{jN}(1-\pi_{jN})}}\left[\frac{N}{mn}\sum_{i=1}^{m}\mathbf{1}_{[0,\pi_{jN})}\left(\frac{R_i-0.5}{N}\right) - \frac{\lceil N\pi_{jN} - 0.5\rceil}{n}\right],$$

where $\lceil\bullet\rceil$ equals the number $\bullet$ when it is an integer, otherwise $\lceil\bullet\rceil = \lfloor\bullet\rfloor + 1$. The above statistic, $\mathcal{L}_j$, contrasts the average values of the rescaled ranks $(R_i - 0.5)/N$, $j = 1, \dots, N$, from the two samples which fall into the interval $[0, \pi_{jN})$. Equivalently, given the vector of ranks $(R_1, \dots, R_N)$, the value of $\mathcal{L}_j$ is a linear function of the values of the rescaled ranks $(R_i - 0.5)/N$, $i = 1, \dots, m$, from the first sample which fall into the interval $[0, \pi_{jN})$. Therefore, it is intuitive that small values of $\mathcal{L}_j$ indicate $\mathbb{H}_1$.

Finally, based on the union-intersection principle, set $M_{\Delta(N)} = \min_{1\le j\le\Delta(N)} \mathcal{L}_j$. For a thorough discussion of the construction and properties of $M_{\Delta(N)}$ in the case where the π_{jN}'s are related to a dyadic partition of (0,1), we refer the reader to Ledwina and Wyłupek [35]. See also Ledwina and Wyłupek [36] for some useful properties of the $\mathcal{L}_j$'s.

Since upper-tailed critical regions have a long tradition in efficiency calculations, for a given $\Delta(N)$, we shall consider

$$\mathcal{T}_N = \max_{1\le j\le\Delta(N)} \{-\mathcal{L}_j\} \tag{3.8}$$

and the related test which rejects $\mathbb{H}_0$ for large values of $\mathcal{T}_N$.

The statistic $\mathcal{T}_N$ differs by some asymptotically negligible quantity from the following weighted Kolmogorov–Smirnov-type statistic:

$$\mathcal{W}_N = \sqrt{\frac{mn}{N}} \max_{1\le j\le\Delta(N)} \frac{(\hat{G}_n - \hat{F}_m)\circ\hat{J}_N^{-1}(\pi_{jN})}{\sqrt{\pi_{jN}(1-\pi_{jN})}}. \tag{3.9}$$

More precisely, the following holds for sufficiently large N:

$$|\mathcal{T}_N - \mathcal{W}_N| \le C_\eta / \sqrt{N \min\{\pi_{1N}, 1-\pi_{\Delta(N)N}\}}, \tag{3.10}$$

where C_η is a positive number which depends only on η. For a proof, see Appendix D.

Our theoretical results are stated under the following basic assumptions: $\Delta(N) = o(N)$ and $1/[\Delta(N)+1] \le \pi_{1N} < \pi_{\Delta(N)N} \le 1 - 1/[\Delta(N)+1]$. Therefore, the right hand side of (3.10) is $o(1)$ and any result that is proven for $\mathcal{T}_N$ is automatically valid for $\mathcal{W}_N$ and *vice versa*. This gives us the flexibility to use the most appropriate and interpretable techniques for particular proofs. For convenience, all of the results are only stated for $\mathcal{T}_N$.

When applying the results of Section 2 to the above statistics, we set $\mathcal{U}_N^{(\mathrm{I})} = \mathcal{V}_N$ and $\mathcal{U}_N^{(\mathrm{II})} = \mathcal{T}_N$.

3.3. Moderate deviations of $\mathcal{V}_N$ and $\mathcal{T}_N$ under P_0^N

Under P_0^N, defined in Section 3.1, one can expect that the tails of $\mathcal{V}_N$ behave similarly to the tails of the classical Kolmogorov–Smirnov statistic for uniformity. Indeed, the ideas developed in Inglot and Ledwina [20] can also be applied to $\mathcal{V}_N$ and we obtain the following result:

Theorem 3.1. *For any real sequence $\{w_N\}$ such that $w_N \to 0$ and $Nw_N^2 \to \infty$, the following holds:*

$$-\lim_{N\to\infty} \frac{1}{Nw_N^2} \log P_0^N\big(\mathcal{V}_N \ge w_N\sqrt{N}\big) = c_{\mathcal{V}} = 2. \tag{3.11}$$

The constant $c_{\mathcal{V}} = 2$, appearing in (3.11), is the index of moderate deviations of $\mathcal{V}$.

To obtain moderate deviations result for $\mathcal{T}_N$, we proceed as follows. From (3.8), $\mathcal{T}_N$ is the maximum of $\Delta(N)$ linear rank statistics with non-continuous score functions, the ℓ_j's. These score functions are simple and can be approximated sufficiently well by piecewise linear functions. Some results proved in Inglot [19] can be applied to the corresponding rank statistics.

Theorem 3.2. *Assume the following: (i) $\Delta(N) \to \infty$, $\Delta(N) = o(N)$, as $N \to \infty$, (ii) $\pi_{1N} \ge 1/[\Delta(N)+1]$, $\pi_{\Delta(N)N} \le 1 - 1/[\Delta(N)+1]$, (iii) $\{w_N\}$ is a sequence of positive numbers such that $w_N \to 0$, $Nw_N^2/\log N \to \infty$, and, for some $\upsilon \in (0,1)$, $w_N^{1-\upsilon}\Delta(N) \to 0$ as $N \to \infty$. Then*

$$-\lim_{N\to\infty} \frac{1}{Nw_N^2} \log P_0^N\big(\mathcal{T}_N \ge w_N\sqrt{N}\big) = c_{\mathcal{T}} = \frac{1}{2}. \tag{3.12}$$

From (3.12), the index of moderate deviations of $\mathcal{T}$ is $c_{\mathcal{T}} = 1/2$.

3.4. The asymptotic behavior of $\mathcal{V}_N$ and $\mathcal{T}_N$ under alternatives satisfying (3.2)

As in Remark 2.2, set $\bar{A}(t) = \bar{A}(t;\eta) = (G_1 - F_1) \circ J_1^{-1}(t)$, for $t \in (0,1)$. Recall also that, under (3.2), $P_{\vartheta_N} \sim F_{1N}$, $Q_{\vartheta_N} \sim G_{1N}$, and $\Pi_{\vartheta_N}^N = P_{\vartheta_N}^{m(N)} \times Q_{\vartheta_N}^{n(N)}$.

First we shall show that (I.2) holds for $\mathcal{U}^{(\mathrm{I})} = \mathcal{V}_N$ with

$$b_{\mathcal{U}^{(\mathrm{I})}}(P_\vartheta \times Q_\vartheta) = \vartheta \sup_{0<t<1} \bar{A}(t;\eta) = \vartheta \sup_{z\in\mathbb{R}} [G_1(z) - F_1(z)].$$

To simplify the notation, we also introduce

$$b_{\mathcal{V}}(\Pi_{\vartheta_N}^N) = \sqrt{\frac{mn}{N}} b_{\mathcal{U}^{(\mathrm{I})}}(P_{\vartheta_N} \times Q_{\vartheta_N}) \tag{3.13}$$

and note that since (F_1, G_1) belongs to $\mathbb{H}_1$, then $b_{\mathcal{V}}(\Pi^N_{\vartheta_N}) > 0$. The following result, along with Lemma B.2 stated in Appendix B, allows us to check whether $\mathcal{V}_N$ satisfies (I.2).

Theorem 3.3. *Assume that $\vartheta_N \in (0,1)$, $\vartheta_N \to 0$ and $N\vartheta_N^2 \to \infty$. Then*

$$\limsup_{N\to\infty} \Pi^N_{\vartheta_N}\big(\mathcal{V}_N - b_{\mathcal{V}}(\Pi^N_{\vartheta_N}) \le w\big) \le V_2(w), \quad w \in \mathbb{R}, \tag{3.14}$$

and

$$\liminf_{N\to\infty} \Pi^N_{\vartheta_N}\big(\mathcal{V}_N - b_{\mathcal{V}}(\Pi^N_{\vartheta_N}) \le w\big) \ge V_1(w), \quad w \in \mathbb{R}_+, \tag{3.15}$$

where $V_1(w) = Pr\big(\sup_{0<t<1} B(t) \le w\big)$, B is a Brownian bridge, $V_2(w) = \Phi(\frac{w}{\sqrt{J_1(z_0)[1-J_1(z_0)]}})$, and Φ denotes the $N(0,1)$ distribution function, while $z_0 = \inf\{z \in \mathbb{R} : G_1(z) - F_1(z) = \sup_{w\in\mathbb{R}}[G_1(w) - F_1(w)]\}$.

Now take a particular sequence $\{\theta_N\}$ such that $\theta_N \in (0,1), \theta_N \to 0$ and $N\theta_N^2 \to \infty$ and the corresponding $\Pi^N_{\theta_N}$. Set

$$b_{\mathcal{T}}(\Pi^N_{\theta_N}) = \theta_N \sqrt{\frac{mn}{N}} \max_{1\le j\le \Delta(N)} \frac{\bar{A}(\pi_{jN})}{\sqrt{\pi_{jN}(1-\pi_{jN})}}.$$

Our next result gives conditions on $\Delta(N)$ and further restrictions on θ_N which guarantee that $\mathcal{U}_N^{(\mathrm{II})} = \mathcal{T}_N$ satisfies (i) and (ii) of Lemma B.1, given in Appendix B.

Theorem 3.4. *Suppose that the following conditions are satisfied: (i) $\pi_{1N} \ge 1/[\Delta(N)+1]$, $\pi_{\Delta(N)N} \le 1 - 1/[\Delta(N)+1]$ and $\max_{1\le j\le\Delta(N)}\{\pi_{jN} - \pi_{j-1N}\} \to 0$, when $N \to \infty$, (ii) $\theta_N \in (0,1)$, $\theta_N \to 0$, $N\theta_N^2/\log^2 N \to \infty$, and $\theta_N\Delta(N) \to 0$, as $N \to \infty$, (iii) $\eta_N \to \eta$, $\eta \in (0,1)$ and $\theta_N(\eta_N - \eta)\sqrt{N} = O(1)$, when $N \to \infty$. Then*

$$\limsup_{N\to\infty} \Pi^N_{\theta_N}\big(\mathcal{T}_N - b_{\mathcal{T}}(\Pi^N_{\theta_N}) \le w\big) \le T_2(w), \quad w \in \mathbb{R}, \tag{3.16}$$

and

$$\liminf_{N\to\infty} \Pi^N_{\theta_N}\big(\mathcal{T}_N - b_{\mathcal{T}}(\Pi^N_{\theta_N}) \le w\big) \ge T_1(w), \quad w \in \mathbb{R}_+, \tag{3.17}$$

where $T_2(w) = \Phi(w)$, while $T_1(w) = Pr(\sup_{t\in[\delta,1-\delta]} |B(t)|/\sqrt{t(1-t)} \le w)$ with B being a Brownian bridge and $\delta = \delta(\bar{A}) \in (0,1)$ being defined by formula (G.3) in Appendix G.

3.5. The main result on the efficiency of $\mathcal{T}_N$ with respect to $\mathcal{V}_N$

The theoretical results presented above allow us to apply Theorem 2.6 and to formulate our basic result on the intermediate efficiency of a class of tests based on $\mathcal{T}_N$.

First, recall that $\bar{A}(t;\eta) = (G_1 - F_1) \circ J_1^{-1}(t)$, $J_1(z) = \eta F_1(z) + (1-\eta)G_1(z)$ and set

$$A^*(t;\eta) = \bar{A}(t;\eta)/\sqrt{t(1-t)}.$$

Theorem 3.5. *Assume that conditions (i) and (iii) of Theorem 3.4 hold and sharpen condition (ii) to (ii)′ $\theta_N \in (0,1)$, $N\theta_N^2/\log^2 N \to \infty$ and $\theta_N^{1-\nu}\Delta(N) \to 0$ for some $\nu \in (0,1)$.*

Then, given $\eta \in (0,1)$, the intermediate efficiency $e_{\mathcal{TV}}$ of $\mathcal{T}$ with respect to $\mathcal{V}$, under $\{P_{\theta_N} \times Q_{\theta_N}\}$, exists and is equal to

$$e_{\mathcal{TV}}(\eta) = \frac{1}{4}\left[\frac{\sup_{0<t<1} A^*(t;\eta)}{\sup_{0<t<1} \bar{A}(t;\eta)}\right]^2 = \frac{1}{4}\left[\frac{\sup_{z\in\mathbb{R}}[(G_1(z)-F_1(z)]/\sqrt{J_1(z)[1-J_1(z)]}}{\sup_{z\in\mathbb{R}}\big[G_1(z)-F_1(z)\big]}\right]^2. \tag{3.18}$$

Besides, for any sequence of alternatives it follows that $e_{\mathcal{TV}}(\eta) \geq 1$ with equality holding if and only if $A^*(t;\eta)$ attains its maximum at $t = 1/2$. Moreover, the choice $\gamma_N = \log N$ and $\lambda_N = N/\Delta(N)^{2/(1-\nu)}$ in (II.1) is adequate for $\mathcal{T}$.

Remark 3.6. It is worth emphasizing that in Theorem 3.5 there are no restrictions on the deviation of the alternative (F_1, G_1) from (F_0, G_0). The assumptions concern only the rates of convergence of θ_N to 0 and $\Delta(N)$ to ∞.

Theorem 3.5 explains qualitatively the outcomes of the simulations in Ledwina and Wyłupek [35, 37]. In Section 4 we demonstrate that the value of $e_{\mathcal{TV}}(\eta)$ also gives precise quantitative information on the relation between the empirical powers of sequences of the statistics $\mathcal{T}$ and $\mathcal{V}$ in the sense described in Remark 2.3.

Remark 3.7. Theorem 3.5 shows how the requirements on the rate of convergence of $\theta_N \to 0$ and $\Delta(N) \to \infty$ can be balanced in order to obtain the efficiency $e_{\mathcal{TV}}(\eta)$. Corollary 3.8 stated below, gives a simple illustration of such relations and their association with the corresponding significance levels of tests. Note also that (i) of Theorem 2.6 and Theorem 3.4 imply $-\log\alpha_N \asymp b^2_{\mathcal{T}}(\Pi^N_{\theta_N})/2 \asymp N\theta_N^2/2$.

Corollary 3.8. *Let $m = \lfloor N\eta \rfloor$, $n = N - m$, $\Delta(N) = \lfloor N^p \rfloor$, $0 < p < 1/2$, and $\theta_N = N^{-q}$ with $0 < p < q < 1/2$. Then the assumptions (i), (ii)' and (iii) of Theorem 3.5 are satisfied for any $\nu < (q-p)/q$ while the corresponding significance levels, for a given q, satisfy $-\log\alpha_N \asymp N^{1-2q}/2$.*

Remark 3.9. The proofs of Theorems 3.2 and 3.4 indicate that $\Delta(N)$ only influences the results by defining the cut-off points from the ends of the interval (0,1). Theorem 3.5 shows that we may consider an interval slightly narrower than $[1/\sqrt{N}, 1-1/\sqrt{N}]$. The choice of the partition points, the π_{jN}'s, inside $[1/(\Delta(N)+1), 1-1/(\Delta(N)+1)]$ is practically immaterial to the final asymptotic results. In practice, it is natural to take a reasonably large number of such points to ensure accuracy, while ensuring that the calculations are not numerically complex. The line of our proofs also makes it clear that, instead of a discretized variant $\mathcal{W}_N$, one may simply consider

$$\sup_{\epsilon(N)\leq t\leq 1-\epsilon(N)} \frac{(\hat{G}_n - \hat{F}_m)\circ \hat{J}_N^{-1}(t)}{\sqrt{t(1-t)}},$$

where $\epsilon(N) \approx o(1/\sqrt{N})$, and obtain similar asymptotic results.

4. Simulation results

We compare the empirical powers of four tests which reject $\mathbb{H}_0$ for large values of the following statistics

$$\mathcal{V}_N = \sqrt{\frac{mn}{N}}\sup_{z\in\mathbb{R}}\Big\{\hat{G}_n(z) - \hat{F}_m(z)\Big\},$$

$$\mathcal{T}_N^\star = \max_{1\leq j\leq\Delta(N)}\big\{-\mathcal{L}_j\big\}, \quad \text{with} \quad \Delta(N) = 2^{\lfloor \log_2 N\rfloor} - 1, \quad \pi_{jN} = \frac{j}{\Delta(N)+1}, \;\; j = 1,\ldots,\Delta(N),$$

$$\mathcal{T}_N^o = \max_{1+\lfloor\sqrt{N}\rfloor\leq j\leq N-\lfloor\sqrt{N}\rfloor}\big\{-\mathcal{L}_j\big\}, \quad \text{where} \;\; \pi_{jN} = \frac{j}{N+1}, \;\; j = 1,\ldots,N,$$

and

$$\mathcal{V}_N^e = \mathcal{V}_{N^{(e)}}, \ \text{ where } N^{(e)} = m^{(e)} + n^{(e)} \ \text{ and } m^{(e)} = \lfloor m e_{\mathcal{TV}} \rfloor, \ n^{(e)} = \lfloor n e_{\mathcal{TV}} \rfloor.$$

$\mathcal{T}_N^\star$ is the variant of $\mathcal{T}_N$ which was studied in Ledwina and Wyłupek [35, 37], in Sections 3.2 and 2.2, respectively. $\mathcal{T}_N^o$ is a new test statistic which, in comparison to $\mathcal{T}_N^\star$, puts less weight on extreme observations. $\mathcal{V}_N^e$ is the two-sample Kolmogorov–Smirnov statistic based on samples of sizes which have been adapted to the intermediate efficiency value $e_{\mathcal{TV}} = e_{\mathcal{TV}}(\eta)$, as defined above.

Taking into account the definition of the intermediate efficiency, it is expected that the empirical power of $\mathcal{V}_N^e = \mathcal{V}_{N^{(e)}}$ will be greater or equal to the corresponding powers of $\mathcal{T}_N^\star$ and $\mathcal{T}_N^o$. We emphasize that in our simulation study, described below, the significance level and alternatives are independent of N. So, although the theory is local and assumes that significance levels tend to 0 as $N \to \infty$, the results demonstrate that this theory accurately describes the exact behavior of power under a standard simulation scheme.

We consider the power and efficiency under some commonly used models of alternatives; *cf.* Barrett and Donald [2], Fan [12], Klonner [32], Schmid and Trede [45]. All of these models were considered in extensive simulation studies in Ledwina and Wyłupek [35, 37]. This enables further comparisons.

A description of the distributions used in our simulation study is given below. We list the models in the same order as they appear in Figures 1–4.

- $MU(a)$, $a \in [-1, 1]$, has density function $1 + a \sin(10\pi x)\mathbf{1}(0.4 < x < 0.6)$, $x \in [0, 1]$; *cf.* Fan [12].
- $Pareto(a)$ coincides with $SM(a, 1, 1)$; see below.
- $SM(a, b, c)$, $a \geq 1$, $b \geq 1$, $c \geq 1$, denotes the Singh–Maddala model which obeys the distribution function $1 - [1 + (x/b)^a]^{-c}$, $x > 0$.
- $LN(a, b)$, $a \in \mathbb{R}$, $b > 0$, is the log-normal distribution with density function $\exp[-(\log x - a)^2/(2b^2)]/(\sqrt{2\pi} bx)$, $x > 0$, while $\beta LN(a_1, b_1) + (1-\beta)LN(a_2, b_2)$, $\beta \in (0, 1)$, denotes a mixture of such distributions.
- $N(a, b)$ denotes the normal distribution with mean a and dispersion b, χ_1^2 denotes the central chi-square distribution with 1 degree of freedom, and $\beta N(a, b) + (1-\beta)\chi_1^2$ is a mixture of such distributions.
- $Laplace(a, b)$, $a \in \mathbb{R}$, $b > 0$, has the density function $\exp(-|x - a|/b)/2b$, $x \in \mathbb{R}$.

The simulation results are presented graphically. The notation used in Figures 1–4 of this paper are identical to that previously used in the two papers by Ledwina and Wyłupek mentioned above. In particular, given the alternative (F_1, G_1), we indicate the underlying distributions in the two samples by F_1/G_1.

The number of Monte Carlo runs is 5000 throughout. The significance level was set to be $\alpha = 0.01$ in all cases apart from the middle row in Figure 4, where it varies over the interval [0.001,0.010].

For each alternative F_1/G_1 under consideration, we plot in the first row of the corresponding figure the values of $e_{\mathcal{TV}} = e_{\mathcal{TV}}(\eta)$ (continuous brown line) and $\arg\max A^*(t;\eta)$ (blue dotted line) against $\eta \in (0, 1)$. Theorem 3.5 shows that the location of $\arg\max A^*(t;\eta)$ on (0,1) is decisive to the magnitude of $e_{\mathcal{TV}}(\eta)$. To calculate the efficiency we had to calculate suprema of functions A^* and $\bar{A}$. These suprema were calculated numerically as a maximum over the set $\{0.0001, 0.0002, \ldots, 0.9999\}$. To calculate values of A^* and $\bar{A}$ in a fixed point we had to calculate a value of function J_1^{-1} in that point. Since the function J_1 is always monotonic this was also calculated numerically using bisection method. The accuracy of calculating J_1^{-1} in a fixed point was 10^{-5}. Moreover, $\arg\max A^*$ was calculated numerically as a point from the set $\{0.0001, 0.0002, \ldots, 0.9999\}$ for which function A^* had the biggest value.

Let us start with a description and some discussion of the results presented in Figures 1–3. For these figures, we took $N = 800$ in the unbalanced case ($\eta_N \neq 0.5$) and various values of $N \in [200, 500]$ for balanced partitions ($\eta_N = 0.5$). The above classification into balanced and unbalanced cases is not very precise, since among the unbalanced cases there is always a balanced one. However, such terminology allows a more succinct description of the results.

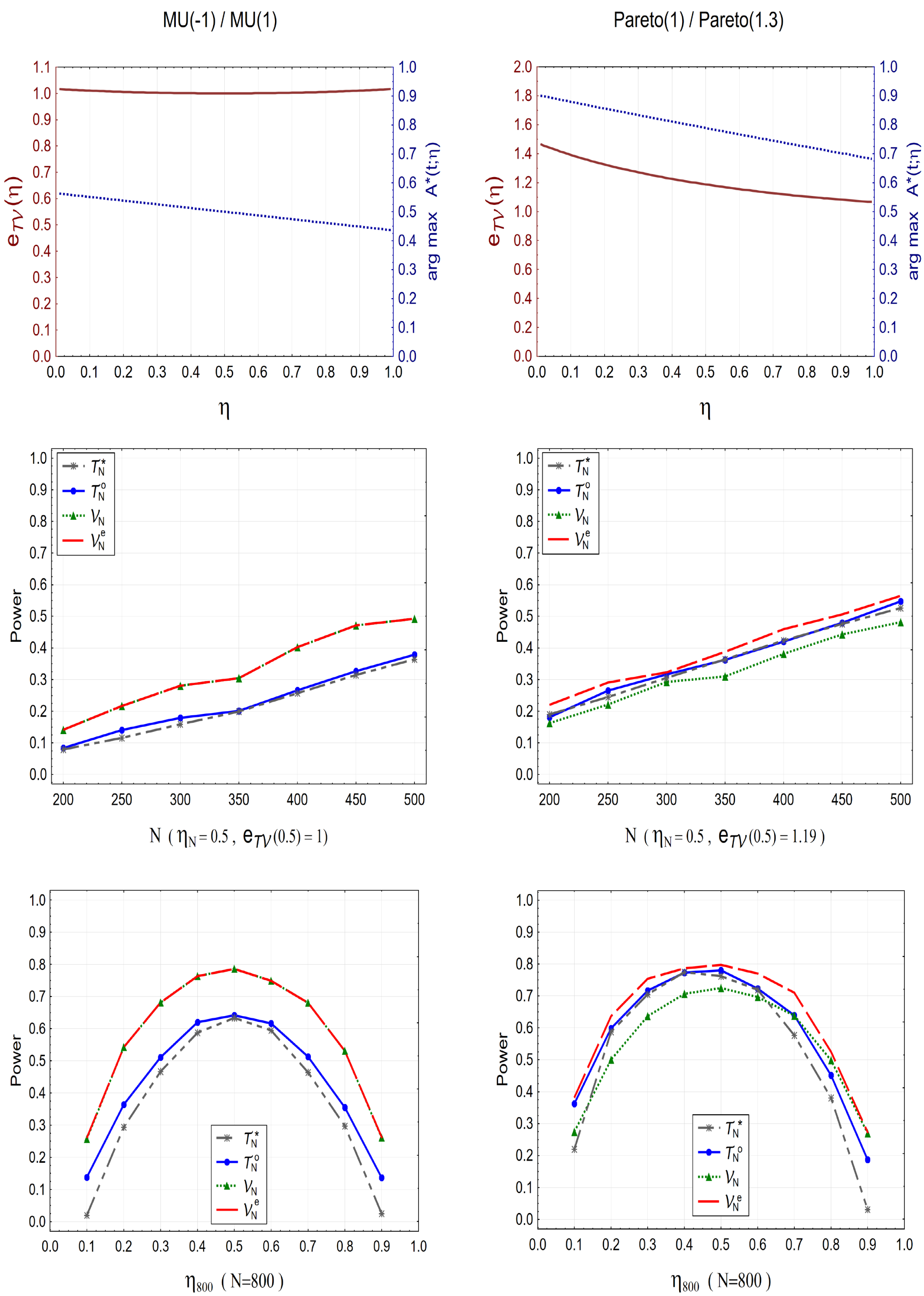

FIGURE 1. Fan and Pareto alternatives. Moderately large sample sizes, $\alpha = 0.01$. *Upper panels*: efficiencies $e_{\mathcal{TV}}(\eta)$ against $\eta \in (0,1)$ – brown continuous line; locations of the maximum of $A^*(t,\eta)$ against $\eta \in (0,1)$ – blue dotted line. *Middle panels*: empirical powers for balanced partitions ($\eta_N = 0.5$) against $N \in [200, 500]$. *Bottom panels*: empirical powers for unbalanced partitions against $\eta_N \in [0.1, 0.9]$; $N = 800$.

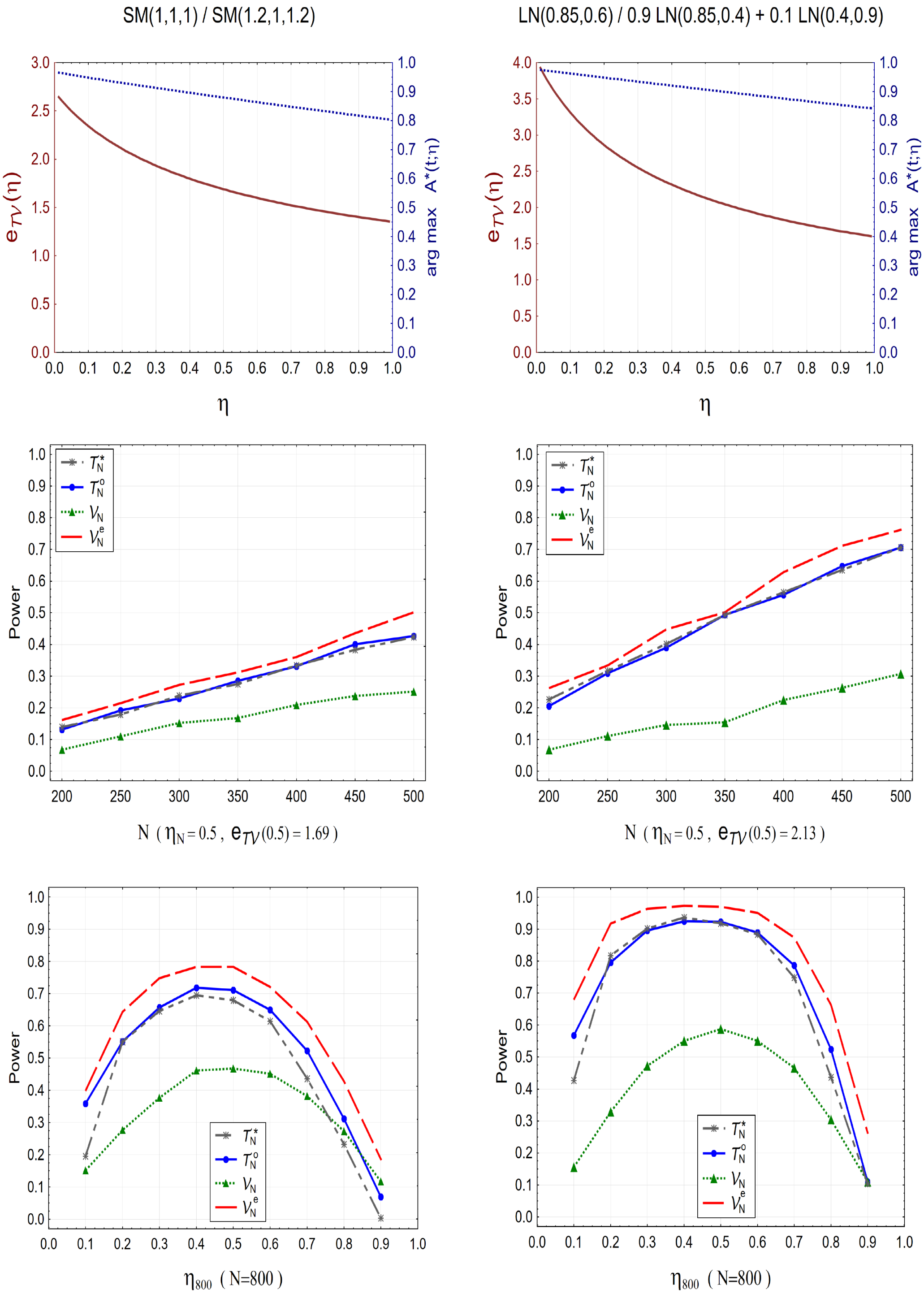

FIGURE 2. Singh–Maddala and log-normal alternatives. Moderately large sample sizes, $\alpha = 0.01$. *Upper panels*: efficiencies $e_{\mathcal{TV}}(\eta)$ against $\eta \in (0,1)$ – brown continuous line; locations of the maximum of $A^*(t,\eta)$ against $\eta \in (0,1)$ – blue dotted line. *Middle panels*: empirical powers for balanced partitions ($\eta_N = 0.5$) against $N \in [200, 500]$. *Bottom panels*: empirical powers for unbalanced partitions against $\eta_N \in [0.1, 0.9]$; $N = 800$.

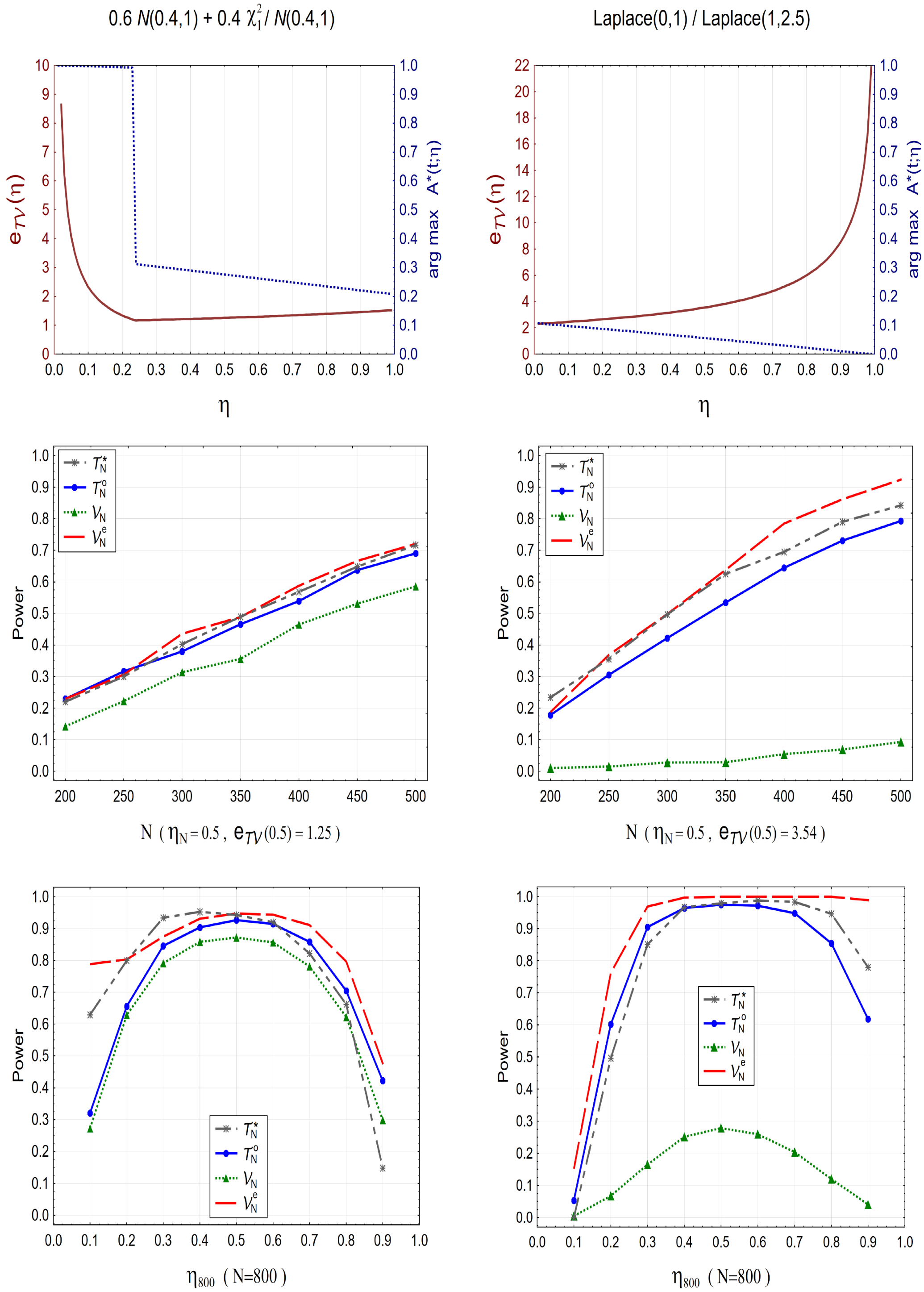

FIGURE 3. Normal-chi-square mixture and Laplace alternatives. Moderately large sample sizes, $\alpha = 0.01$. *Upper panels*: efficiencies $e_{\mathcal{TV}}(\eta)$ against $\eta \in (0,1)$ – brown continuous line; locations of the maximum of $A^*(t,\eta)$ against $\eta \in (0,1)$ – blue dotted line. *Middle panels*: empirical powers for balanced partitions ($\eta_N = 0.5$) against $N \in [200, 500]$. *Bottom panels*: empirical powers for unbalanced partitions against $\eta_N \in [0.1, 0.9]$; $N = 800$.

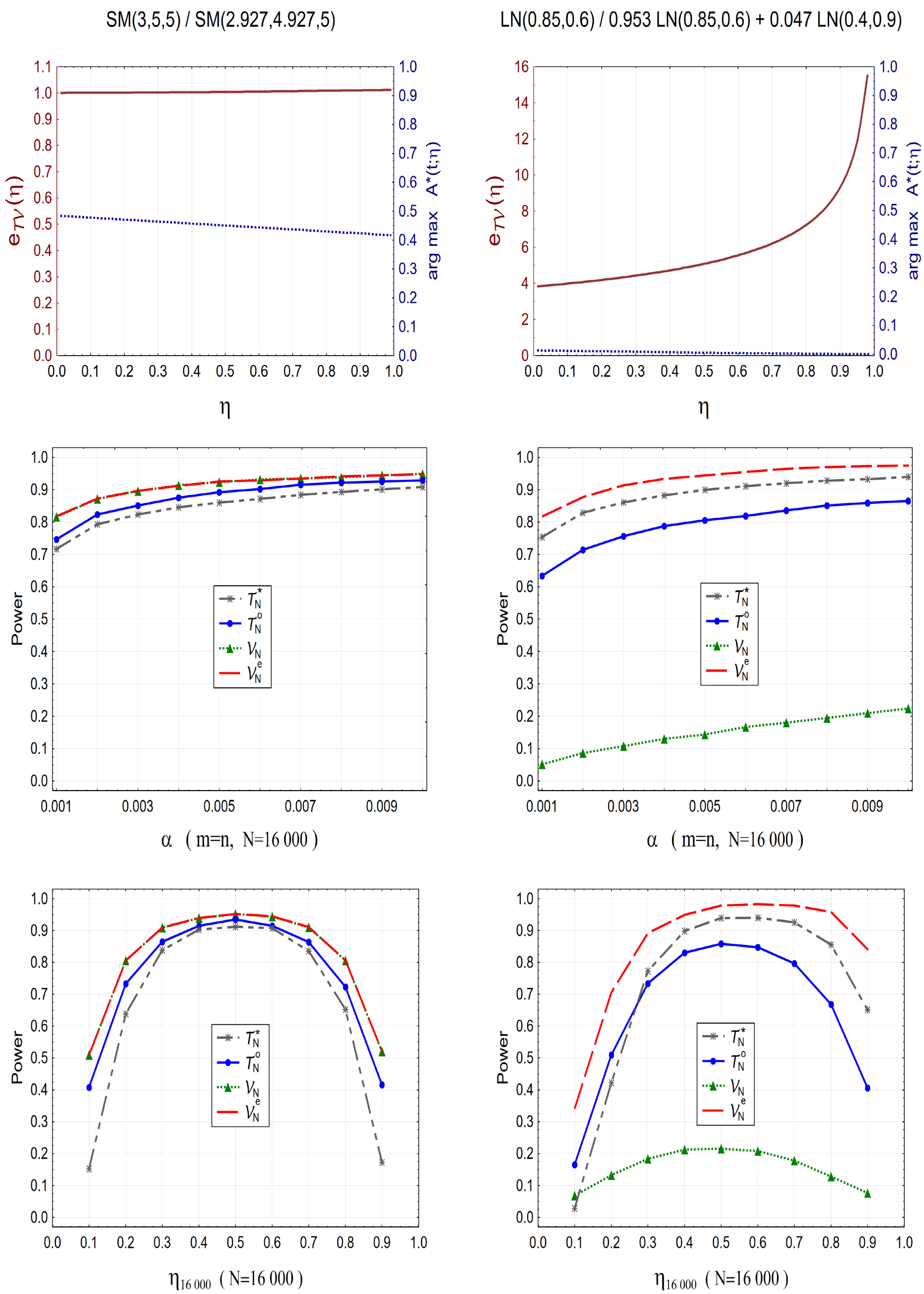

FIGURE 4. Singh–Maddala and log-normal alternatives. Large sample sizes. *Upper panels*: efficiencies $e_{\mathcal{TV}}(\eta)$ against $\eta \in (0,1)$ – brown continuous line; locations of the maximum of $A^*(t,\eta)$ against $\eta \in (0,1)$ – blue dotted line. *Middle panels*: empirical powers for balanced partitions ($\eta_N = 0.5$) against $\alpha \in [0.001, 0.010]$; $N = 16\ 000$. *Bottom panels*: empirical powers for unbalanced partitions against η_N; $N = 16\ 000$; $\alpha = 0.01$.

Figures 1–3 are ordered such that the maximal attainable value of $e_{\mathcal{TV}}(\eta)$ over $\eta \in (0,1)$ is increasing. In Figures 1–3, the range of η is $[0.01, 0.99]$. Observe that the maximal value of $e_{\mathcal{TV}}(\eta)$ over the above mentioned interval ranges from 1.016, in the case of the alternative $MU(-1)/MU(1)$, to 22, for the alternative $Laplace(0,1)/Laplace(1,1.25)$. Since the vertical scales in the first rows of Figures 1–3 are different in each case, in order to increase readability, we display the value of $e_{\mathcal{TV}}(0.5)$ in the middle row of each figure.

In Figure 1, we illustrate the empirical powers of the four tests for the selection of sample sizes described above, both for balanced and unbalanced partitions, under two pairs of alternative distributions, corresponding to the Fan and Pareto models, respectively. The middle row contains the results for balanced partitions, while the bottom row describes the results for unbalanced ones. Figures 2 and 3 are constructed in an analogous manner.

The behavior of the test based on $\mathcal{V}_N^e$ illustrates how the intuitive meaning of the efficiency measure manifests itself for finite sample sizes. It is expected that when $e_{\mathcal{TV}} > 1$, the empirical powers of $\mathcal{V}_N^e$ should be slightly higher than the corresponding powers of $\mathcal{T}_N^\star$ and $\mathcal{T}_N^o$. We see that indeed this is the case regardless of the model, exact form of the efficiency function and whether the partition is balanced or not. We considered $\eta_N \in [0.1, 0.9]$. For moderately large sample sizes, as used in the cases presented in Figures 1–3, the accuracy of the prediction of the empirical power of $\mathcal{V}_N^e$ is very high for $\eta_N \in [0.2, 0.8]$.

In the cases where $e_{\mathcal{TV}}(\eta)$ is 1 or very close to 1 for all η (*cf.* Fig. 1), the empirical powers of $\mathcal{V}_N$ may be greater than the corresponding powers of $\mathcal{T}_N^\star$ and $\mathcal{T}_N^o$. The alternative that we consider in the first column of Figure 1, based on Fan [12], may serve as an illustration of such a situation. In a sense, this is the least favorable situation for weighted statistics which are designed to be sensitive to differences between tails. Indeed, the two distributions F_1 and G_1 have the same tails and differ only in the central part. In such a case, the simpler structure of $\mathcal{V}_N$ plays a role. It can also be seen from Figure 1 that in a slightly less extreme case, represented by the pair of Pareto distributions, this deficiency disappears.

The evidence provided in Figures 1–3 clearly shows that, except in some very difficult circumstances (very small or very large η_N), applying the concept of the intermediate efficiency to finite samples works very well. This situation should obviously be even better when N is larger; *cf.* Figure 4 and the related comments given below.

We also studied larger sample sizes and smaller significance levels. In Figure 4, we display the results of such experiments for two models: one, corresponding to Singh–Maddala distributions, where the efficiency is very close to 1 for all η, and the second one, corresponding to log-normal distributions, where the efficiency lies in an interval approximately equal to [4,16], depending on the value of η. In the middle row, we present empirical powers against $\alpha \in [0.001, 0.010]$ for balanced partitions and $N = 16\,000$. We see that the efficiency gives very accurate and stable results. The same comment is valid for empirical powers under unbalanced partitions and $\alpha = 0.01$. These results are presented in the bottom panels of Figure 4. It can be seen that for all $\eta \in [0.1, 0.9]$ the simulation results are satisfactory.

To close, we comment on the differences between the empirical behavior of $\mathcal{T}_N^\star$ and $\mathcal{T}_N^o$. Our simulations show that $\mathcal{T}_N^\star$ and $\mathcal{T}_N^o$ behave similarly for balanced partitions as long as the efficiency $e_{\mathcal{TV}}$ is not very large. In the opposite case (*cf.* Figs. 2–4), there is some gain from using $\mathcal{T}_N^\star$. The explanation of this is simple. High efficiency occurs when arg max $A^\star(t;\eta)$ is located near 0 or 1. Obviously $\mathcal{T}_N^\star$ has a greater chance to detect such changes, as it allows closer inspection of weighted two-sample rank processes at arguments which are closer to 0 and 1 than $\mathcal{T}_N^o$ does; *cf.* (3.9) and related comments. In the cases of highly unbalanced partitions and small or moderate efficiencies, the new solution $\mathcal{T}_N^o$ provides some improvement under most of the alternatives. It is also worth noticing that the value of $\Delta(N)$ corresponding to $\mathcal{T}_N^o$ approximately satisfies the requirements of our theoretical results. For $\mathcal{T}_N^\star$, the choice of $\Delta(N)$ is outside the allowable range. In any case, applying the concept of intermediate efficiency to finite samples works well for $\mathcal{T}_N^\star$ in our simulation experiments.

Appendix A. Proof of Theorem 2.6

Step 1. ***Proof of (i) and (ii).*** Let $\{\alpha_N\}$ be a sequence from $\mathbb{L}^*$ satisfying (2.11). For any $h > 0$ define $w_N = \sqrt{(-h \log \alpha_N)/N c_{\mathcal{U}^{(\mathrm{II})}}}$. For this w_N (2.9) applies and yields

$$\frac{c_{\mathcal{U}^{(\mathrm{II})}}}{h \log \alpha_N} \log \sup_{P\times Q\in\mathbb{P}_0} P^{m(N)} \times Q^{n(N)}\big(\mathcal{U}_N^{(\mathrm{II})} \geq w_N\sqrt{N}\big) \to c_{\mathcal{U}^{(\mathrm{II})}}. \tag{A.1}$$

For arbitrary $\epsilon > 0$ take $h = 1+\epsilon$ in w_N appearing in (A.1). Then, for sufficiently large N,

$$\sup_{P\times Q\in\mathbb{P}_0} P^{m(N)} \times Q^{n(N)}\big(\mathcal{U}_N^{(\mathrm{II})} \geq w_N\sqrt{N}\big) \leq \alpha_N$$

which, by the definition of $u^{(\mathrm{II})}_{\alpha N}$, means that $u^{(\mathrm{II})}_{\alpha_N N} \leq \sqrt{(-(1+\epsilon)\log\alpha_N)/c_{\mathcal{U}^{(\mathrm{II})}}}$ or equivalently $c_{\mathcal{U}^{(\mathrm{II})}}[u^{(\mathrm{II})}_{\alpha_N N}]^2 \leq -(1+\epsilon)\log\alpha_N$. Similarly taking in (A.1) $h = 1-\epsilon$ we get $c_{\mathcal{U}^{(\mathrm{II})}}[u^{(\mathrm{II})}_{\alpha_N N}]^2 \geq -(1-\epsilon)\log\alpha_N$ for sufficiently large N. Since ϵ was taken arbitrarily we obtain

$$-\log\alpha_N = c_{\mathcal{U}^{(\mathrm{II})}}[u^{(\mathrm{II})}_{\alpha_N N}]^2[1+o(1)]. \tag{A.2}$$

On the other hand, by (2.10), we have for arbitrary $\epsilon > 0$

$$\Pi^N_{\theta_N}\big(\mathcal{U}_N^{(\mathrm{II})} \geq (1+\epsilon)b_{\mathcal{U}^{(\mathrm{II})}}(\Pi^N_{\theta_N})\big) \to 0, \ \ \Pi^N_{\theta_N}\big(\mathcal{U}_N^{(\mathrm{II})} \geq (1-\epsilon)b_{\mathcal{U}^{(\mathrm{II})}}(\Pi^N_{\theta_N})\big) \to 1.$$

Since $\{\alpha_N\} \in \mathbb{L}^*$, the condition (2.4) implies $(1-\epsilon)b_{\mathcal{U}^{(\mathrm{II})}}(\Pi^N_{\theta_N}) \leq u^{(\mathrm{II})}_{\alpha_N N} \leq (1+\epsilon)b_{\mathcal{U}^{(\mathrm{II})}}(\Pi^N_{\theta_N})$ for sufficiently large N. As ϵ is arbitrary this means that $u^{(\mathrm{II})}_{\alpha_N N} = b_{\mathcal{U}^{(\mathrm{II})}}(\Pi^N_{\theta_N})[1+o(1)]$ which together with (A.2) gives

$$-\log\alpha_N = c_{\mathcal{U}^{(\mathrm{II})}}\big[b_{\mathcal{U}^{(\mathrm{II})}}(\Pi^N_{\theta_N})\big]^2[1+o(1)], \tag{A.3}$$

and proves (i).

Now we will prove (ii). Let $\{\alpha'_N\}$ be any sequence from $\mathbb{L}^*$. Then

$$0 < \liminf_{N\to\infty} \Pi^N_{\theta_N}\big(\mathcal{U}_N^{(\mathrm{II})} > u^{(\mathrm{II})}_{\alpha'_N N}\big) \leq \limsup_{N\to\infty} \Pi^N_{\theta_N}\big(\mathcal{U}_N^{(\mathrm{II})} > u^{(\mathrm{II})}_{\alpha'_N N}\big) < 1.$$

Hence, for N large enough, we have

$$C_1 < \Pi^N_{\theta_N}\left(\frac{\mathcal{U}_N^{(\mathrm{II})}}{b_{\mathcal{U}^{(\mathrm{II})}}(\Pi^N_{\theta_N})} - 1 > \frac{u^{(\mathrm{II})}_{\alpha'_N N}}{b_{\mathcal{U}^{(\mathrm{II})}}(\Pi^N_{\theta_N})} - 1\right) < C_2,$$

for some constants $0 < C_1 < C_2 < 1$. On the other hand, using (II.2) we obtain for every $\epsilon > 0$ and for large N

$$-\epsilon < \frac{u^{(\mathrm{II})}_{\alpha'_N N}}{b_{\mathcal{U}^{(\mathrm{II})}}(\Pi^N_{\theta_N})} - 1 < \epsilon.$$

The above yields

$$u^{(\mathrm{II})}_{\alpha'_N N} = \big[b_{\mathcal{U}^{(\mathrm{II})}}(\Pi^N_{\theta_N})\big]\,[1+o(1)].$$

Using (A.3) and the above equation we obtain

$$u^{(\mathrm{II})}_{\alpha'_N N} = \sqrt{\frac{-\log\alpha_N}{c_{\mathcal{U}^{(\mathrm{II})}}}}(1+o(1)).$$

Since $\{\alpha_N\}$ satisfy (2.11) then for $w_N = hu^{(\mathrm{II})}_{\alpha'_N N}/\sqrt{N}$, where $h>0$ is any constant, (2.9) applies and we have

$$\lim_{N\to\infty}\frac{c_{\mathcal{U}^{(\mathrm{II})}}}{h^2\log\alpha_N}\log\sup_{P\times Q\in\mathbb{P}_0} P^{m(N)}\times Q^{n(N)}\big(\mathcal{U}^{(\mathrm{II})}_N > hu^{(\mathrm{II})}_{\alpha'_N N}\big) = c_{\mathcal{U}^{(\mathrm{II})}}.$$

For any $\epsilon>0$, from the definition of $u^{(\mathrm{II})}_{\alpha'_N N}$ it holds that

$$\sup_{P\times Q\in\mathbb{P}_0} P^{m(N)}\times Q^{n(N)}\big(\mathcal{U}^{(\mathrm{II})}_N > (1+\epsilon)u^{(\mathrm{II})}_{\alpha'_N N}\big)\le \alpha'_N,$$

$$\sup_{P\times Q\in\mathbb{P}_0} P^{m(N)}\times Q^{n(N)}\big(\mathcal{U}^{(\mathrm{II})}_N > (1-\epsilon)u^{(\mathrm{II})}_{\alpha'_N N}\big)\ge \alpha'_N.$$

Hence, by taking $h=1+\epsilon$ and $h=1-\epsilon$ in the previous equality, we obtain

$$(1-\epsilon)^2\le\liminf_{N\to\infty}\frac{\log\alpha'_N}{\log\alpha_N}\le\limsup_{N\to\infty}\frac{\log\alpha'_N}{\log\alpha_N}\le(1+\epsilon)^2,$$

which proves (ii).

Finally note that, from (2.12) and (A.3) it follows that

$$-\log\alpha_N = \mathbf{e}c_{\mathcal{U}^{(\mathrm{I})}}\frac{m(N)n(N)}{N}\big[b_{\mathcal{U}^{(\mathrm{I})}}(P_{\theta_N}\times Q_{\theta_N})\big]^2[1+o(1)]\quad\text{if }\ \mathbf{e}\in(0,\infty), \tag{A.4}$$

$$-\log\alpha_N = o\left(\frac{mn}{N}\big[b_{\mathcal{U}^{(\mathrm{I})}}(P_{\theta_N}\times Q_{\theta_N})\big]^2\right)\quad\text{if }\ \mathbf{e}=0, \tag{A.5}$$

$$\frac{mn}{N}\big[b_{\mathcal{U}^{(\mathrm{I})}}(P_{\theta_N}\times Q_{\theta_N})\big]^2 = o(-\log\alpha_N)\quad\text{if }\ \mathbf{e}=\infty. \tag{A.6}$$

The statements (A.4)–(A.6) are also true for any $\{\alpha'_N\}\in\mathbb{L}^*$ since $\log\alpha_N/\log\alpha'_N\to 1$.

Step 2. ***Lower bound for the fraction of sample sizes.*** For $\mathbf{e}\in(0,\infty]$ we shall show that (*cf.* (2.7))

$$\liminf_{N\to\infty}\frac{M_{\mathcal{U}^{(\mathrm{II})}\mathcal{U}^{(\mathrm{I})}}(N,\pi)}{N}\ge\mathbf{e}. \tag{A.7}$$

Suppose, contrary, that there exists an increasing sequence $\{k_j\}$ of natural numbers such that $k_j\to\infty$ as $j\to\infty$ and

$$\frac{M_{\mathcal{U}^{(\mathrm{II})}\mathcal{U}^{(\mathrm{I})}}(k_j,\pi)}{k_j}\to\gamma<\mathbf{e}.$$

For $\tau > 0$ such that $\gamma + \tau < \mathbf{e}$, define

$$g_j = \frac{M_{\mathcal{U}^{(\mathrm{II})}\mathcal{U}^{(\mathrm{I})}}(k_j, \pi)}{k_j} + \frac{\lfloor \tau k_j \rfloor}{k_j}.$$

Then $\lim_{j\to\infty} g_j = \gamma + \tau \in (0, \mathbf{e})$. Moreover, $\{g_j k_j\}$ is a sequence of integers and $g_j k_j \ge M_{\mathcal{U}^{(\mathrm{II})}\mathcal{U}^{(\mathrm{I})}}(k_j, \pi)$ for sufficiently large j. Hence by (2.6)

$$\Pi^{g_j k_j}_{\theta_{k_j}} (\mathcal{U}^{(\mathrm{I})}_{g_j k_j} > u^{(\mathrm{I})}_{\alpha_{k_j} g_j k_j}) \ge \Pi^{k_j}_{\theta_{k_j}} (\mathcal{U}^{(\mathrm{II})}_{k_j} > u^{(\mathrm{II})}_{\alpha_{k_j} k_j}). \tag{A.8}$$

Since $\{g_j\}$ has positive and finite limit and $\{\alpha_N\} \in \mathbb{L}^* \subset \mathbb{L}$ the assumption (I.1) can be applied to the sequence $\{w_N\}$ defined as follows: $w^2_{g_j k_j} = -(1-\delta)(\log \alpha_{k_j})/g_j k_j c_{\mathcal{U}^{(\mathrm{I})}}$ for $j = 1, 2, \dots$ and $w^2_N = -(1-\delta)(\log \alpha_N)/N c_{\mathcal{U}^{(\mathrm{I})}}$ for $N \neq g_j k_j$, where $\delta \in (0,1)$ is arbitrary. The assumption (I.1) applied to the subsequence $\{g_j k_j\}$ yields

$$\frac{c_{\mathcal{U}^{(\mathrm{I})}}}{(1-\delta)\log \alpha_{k_j}} \log \sup_{P\times Q \in \mathbb{P}_0} P^{m(g_j k_j)} \times Q^{n(g_j k_j)} \left(\mathcal{U}^{(\mathrm{I})}_{g_j k_j} \ge \sqrt{\frac{-(1-\delta)\log \alpha_{k_j}}{c_{\mathcal{U}^{(\mathrm{I})}}}} \right) \to c_{\mathcal{U}^{(\mathrm{I})}},$$

which means that for sufficiently large j

$$\sup_{P\times Q \in \mathbb{P}_0} P^{m(g_j k_j)} \times Q^{n(g_j k_j)} \left(\mathcal{U}^{(\mathrm{I})}_{g_j k_j} \ge \sqrt{\frac{-(1-\delta)\log \alpha_{k_j}}{c_{\mathcal{U}^{(\mathrm{I})}}}} \right) \ge \alpha_{k_j}.$$

This and the definition of $u^{(\mathrm{I})}_{\alpha N}$ imply

$$u^{(\mathrm{I})}_{\alpha_{k_j} g_j k_j} \ge \sqrt{\frac{-(1-\delta)\log \alpha_{k_j}}{c_{\mathcal{U}^{(\mathrm{I})}}}}.$$

Hence, from (A.8) and the fact that $\{\alpha_N\} \in \mathbb{L}^*$ we obtain

$$\liminf_{j\to\infty} \Pi^{g_j k_j}_{\theta_{k_j}} \left(\mathcal{U}^{(\mathrm{I})}_{g_j k_j} \ge \sqrt{\frac{-(1-\delta)\log \alpha_{k_j}}{c_{\mathcal{U}^{(\mathrm{I})}}}} \right) > 0. \tag{A.9}$$

Now, consider a sequence $\{\vartheta_N\}$, being a modification of $\{\theta_N\}$, and defined as follows: $\vartheta_{g_j k_j} = \theta_{k_j}$ for $j = 1, 2, \dots$ and $\vartheta_N = \theta_N$ for $N \neq g_j k_j$.

Since $g_j \to \gamma + \tau \in (0, \mathbf{e})$ we have $\vartheta_N \to 0$ and $N\vartheta^\rho_N \to \infty$ and (I.2) can be used for this sequence. Hence (2.8) applied to the subsequence $\{g_j k_j\}$ and arbitrary $\epsilon > 0$ implies

$$\Pi^{g_j k_j}_{\theta_{k_j}} \left(\mathcal{U}^{(\mathrm{I})}_{g_j k_j} \ge (1+\epsilon)\sqrt{\frac{m(g_j k_j) n(g_j k_j)}{g_j k_j}} b_{\mathcal{U}^{(\mathrm{I})}}(P_{\theta_{k_j}} \times Q_{\theta_{k_j}}) \right) \to 0. \tag{A.10}$$

We shall show that the relations (A.9) and (A.10) give a contradiction.

Assume first that $\mathbf{e} \in (0, \infty)$. We have $\gamma + \tau < \mathbf{e}$. For fixed $\delta < 1 - (\gamma+\tau)/\mathbf{e}$ choose $\epsilon > 0$ so small that $(1+\epsilon)^2(\gamma+\tau) < (1-\delta)\mathbf{e}$. Define $\kappa_N = \sqrt{\eta_N(1-\eta_N)} = \sqrt{m(N)n(N)}/N$. By (A.4), the convergence $\kappa_N \to \kappa$

and $g_j \to \gamma + \tau$, and the choice of δ and ϵ we have, for j sufficiently large,

$$\begin{aligned}\frac{-(1-\delta)\log\alpha_{k_j}}{c_{\mathcal{U}^{(\mathrm{I})}}} &= (1-\delta)\mathbf{e}k_j\kappa_{k_j}^2\big[b_{\mathcal{U}^{(\mathrm{I})}}(P_{\theta_{k_j}}\times Q_{\theta_{k_j}})\big]^2[1+o(1)]\\ &= (1-\delta)\frac{\mathbf{e}}{\gamma+\tau}\frac{\kappa_{k_j}^2}{\kappa_{g_jk_j}^2}\frac{m(g_jk_j)n(g_jk_j)}{g_jk_j}\big[b_{\mathcal{U}^{(\mathrm{I})}}(P_{\theta_{k_j}}\times Q_{\theta_{k_j}})\big]^2[1+o(1)]\\ &> (1+\epsilon)^2\frac{m(g_jk_j)n(g_jk_j)}{g_jk_j}\big[b_{\mathcal{U}^{(\mathrm{I})}}(P_{\theta_{k_j}}\times Q_{\theta_{k_j}})\big]^2\end{aligned}$$

which contradicts (A.9) and (A.10).

If $\mathbf{e} = \infty$ we have from (A.6) and the convergence $g_j \to \gamma + \tau$ and $\kappa_N \to \kappa = \sqrt{\eta(1-\eta)}$

$$\begin{aligned}\frac{m(g_jk_j)n(g_jk_j)}{g_jk_j}\big[b_{\mathcal{U}^{(\mathrm{I})}}(P_{\theta_{k_j}}\times Q_{\theta_{k_j}})\big]^2 &= \frac{\kappa_{g_jk_j}^2}{\kappa_{k_j}^2}g_j\frac{m(k_j)n(k_j)}{k_j}\big[b_{\mathcal{U}^{(\mathrm{I})}}(P_{\theta_{k_j}}\times Q_{\theta_{k_j}})\big]^2\\ &= o\Big(\frac{-(1-\delta)\log\alpha_{k_j}}{c_{\mathcal{U}^{(\mathrm{I})}}}\Big)\end{aligned}$$

which contradicts (A.9) and (A.10), as well.

Step 3. ***Upper bound for the fraction of sample sizes.*** For $\mathbf{e} \in [0,\infty)$ we shall show that

$$\limsup_{N\to\infty}\frac{M_{\mathcal{U}^{(\mathrm{II})}\mathcal{U}^{(\mathrm{I})}}(N,\pi)}{N} \le \mathbf{e}. \tag{A.11}$$

The argument is very similar to that of Step 2. Suppose, that there exists an increasing sequence $\{k_j\}$ of natural numbers such that

$$g_j = \frac{M_{\mathcal{U}^{(\mathrm{II})}\mathcal{U}^{(\mathrm{I})}}(k_j,\pi)-1}{k_j} \to \gamma > \mathbf{e}.$$

Note that γ may be equal to ∞. Since $g_jk_j = M_{\mathcal{U}^{(\mathrm{II})}\mathcal{U}^{(\mathrm{I})}}(k_j,\pi) - 1$ then by (2.6)

$$\Pi_{\theta_{k_j}}^{g_jk_j}\big(\mathcal{U}_{g_jk_j}^{(\mathrm{I})} > u_{\alpha_{k_j}g_jk_j}^{(\mathrm{I})}\big) < \Pi_{\theta_{k_j}}^{k_j}\big(\mathcal{U}_{g_jk_j}^{(\mathrm{II})} > u_{\alpha_{k_j}k_j}^{(\mathrm{II})}\big). \tag{A.12}$$

Since $\{g_j\}$ has positive limit or tends to ∞ and $\{\alpha_N\} \in \mathbb{L}^* \subset \mathbb{L}$, then the condition (I.1) can be applied to the sequence $\{w_N\}$ defined as follows: $w_{g_jk_j}^2 = -(1+\delta)(\log\alpha_{k_j})/g_jk_jc_{\mathcal{U}^{(\mathrm{I})}}$ for $j = 1,2,\ldots$ and $w_N^2 = -(1+\delta)(\log\alpha_N)/Nc_{\mathcal{U}^{(\mathrm{I})}}$ for $N \ne g_jk_j$, where $\delta > 0$ is arbitrary. By (I.1) applied to the subsequence $\{g_jk_j\}$ we get

$$\frac{c_{\mathcal{U}^{(\mathrm{I})}}}{(1+\delta)\log\alpha_{k_j}}\log\sup_{P\times Q\in\mathbb{P}_0}P^{m(g_jk_j)}\times Q^{n(g_jk_j)}\left(\mathcal{U}_{g_jk_j}^{(\mathrm{I})} \ge \sqrt{\frac{-(1+\delta)\log\alpha_{k_j}}{c_{\mathcal{U}^{(\mathrm{I})}}}}\right) \to c_{\mathcal{U}^{(\mathrm{I})}}$$

which means that for sufficiently large j

$$\sup_{P\times Q\in\mathbb{P}_0}P^{m(g_jk_j)}\times Q^{n(g_jk_j)}\left(\mathcal{U}_{g_jk_j}^{(\mathrm{I})} \ge \sqrt{\frac{-(1+\delta)\log\alpha_{k_j}}{c_{\mathcal{U}^{(\mathrm{I})}}}}\right) \le \alpha_{k_j}.$$

This and the definition of $u^{(\mathrm{I})}_{\alpha N}$ imply

$$u^{(\mathrm{I})}_{\alpha_{k_j} g_j k_j} \leq \sqrt{\frac{-(1+\delta)\log \alpha_{k_j}}{c_{\mathcal{U}^{(\mathrm{I})}}}}.$$

Hence, from (A.12) and the fact that $\{\alpha_N\} \in \mathbb{L}^*$ we obtain

$$\limsup_{j\to\infty} \Pi^{g_j k_j}_{\theta_{k_j}} \left(\mathcal{U}^{(\mathrm{I})}_{g_j k_j} \geq \sqrt{\frac{-(1+\delta)\log \alpha_{k_j}}{c_{\mathcal{U}^{(\mathrm{I})}}}} \right) < 1. \tag{A.13}$$

Now, consider a sequence $\{\vartheta_N\}$, being a modification of $\{\theta_N\}$, which is defined as follows: $\vartheta_{g_j k_j} = \theta_{k_j}$ for $j = 1, 2, \ldots$ and $\vartheta_N = \theta_N$ for $N \neq g_j k_j$, where $\{k_j\}$ is the sequence selected at the beginning of this step.

Since $\gamma \in (\mathbf{e}, \infty]$ we have $\vartheta_N \to 0$ and $N\vartheta_N^{\rho} \to \infty$ and (I.2) can be applied to this sequence. Hence (2.8) applied for the subsequence $\{g_j k_j\}$ and arbitrary $\epsilon > 0$ yields

$$\Pi^{g_j k_j}_{\theta_{k_j}} \left(\mathcal{U}^{(\mathrm{I})}_{g_j k_j} \geq (1-\epsilon) \sqrt{\frac{m(g_j k_j) n(g_j k_j)}{g_j k_j}} b_{\mathcal{U}^{(\mathrm{I})}}(P_{\theta_{k_j}} \times Q_{\theta_{k_j}}) \right) \to 1. \tag{A.14}$$

We shall argue that the relations (A.13) and (A.14) give a contradiction.

Indeed, if $\mathbf{e} \in (0, \infty)$ we have $\gamma > \mathbf{e}$. For fixed $\delta < \gamma/\mathbf{e} - 1$ choose $\epsilon > 0$ so small that $(1-\epsilon)^2 \gamma > (1+\delta)\mathbf{e}$. This, (A.4), $\kappa_N \to \kappa$ and $g_j \to \gamma$ imply for j sufficiently large

$$\begin{aligned}
\frac{-(1+\delta)\log \alpha_{k_j}}{c_{\mathcal{U}^{(\mathrm{I})}}} &= (1+\delta)\mathbf{e} k_j \kappa^2_{k_j} [b_{\mathcal{U}^{(\mathrm{I})}}(P_{\theta_{k_j}} \times Q_{\theta_{k_j}})]^2 [1 + o(1)] \\
&= (1+\delta) \frac{\mathbf{e}}{\gamma} \frac{\kappa^2_{k_j}}{\kappa^2_{g_j k_j}} \frac{m(g_j k_j) n(g_j k_j)}{g_j k_j} \left[b_{\mathcal{U}^{(\mathrm{I})}}(P_{\theta_{k_j}} \times Q_{\theta_{k_j}}) \right]^2 [1 + o(1)] \\
&< (1-\epsilon)^2 \frac{m(g_j k_j) n(g_j k_j)}{g_j k_j} \left[b_{\mathcal{U}^{(\mathrm{I})}}(P_{\theta_{k_j}} \times Q_{\theta_{k_j}}) \right]^2
\end{aligned}$$

which contradicts (A.13) and (A.14).

If $\mathbf{e} = 0$ we have from (A.5) and the convergence $g_j \to \gamma > 0$ and $\kappa_N \to \kappa$

$$\begin{aligned}
\frac{-(1+\delta)\log \alpha_{k_j}}{c_{\mathcal{U}^{(\mathrm{I})}}} &= o\Big(k_j \left[b_{\mathcal{U}^{(\mathrm{I})}}(P_{\theta_{k_j}} \times Q_{\theta_{k_j}}) \right]^2 \Big) \\
&= o\left(\frac{1}{\kappa^2_{g_j k_j}} \frac{1}{g_j} \frac{m(g_j k_j) n(g_j k_j)}{g_j k_j} \left[b_{\mathcal{U}^{(\mathrm{I})}}(P_{\theta_{k_j}} \times Q_{\theta_{k_j}}) \right]^2 \right)
\end{aligned}$$

which contradicts (A.13) and (A.14), as well. The proof is complete. □

Appendix B. Lemma B.1, Lemma B.2, and proof of Lemma B.1

Lemma B.1. *Let $\{P_{\theta_N} \times Q_{\theta_N}\}$ be the particular sequence of alternatives under consideration. Suppose that there exist cumulative distribution functions $U_1^{(\mathrm{II})}$ and $U_2^{(\mathrm{II})}$ and positive sequences $\{a_N^{(\mathrm{II})}\}$ and $\{b_N^{(\mathrm{II})}\}$ such that $b_N^{(\mathrm{II})} \to \infty$, $b_N^{(\mathrm{II})}/a_N^{(\mathrm{II})} \to \infty$ and for some $w_0^{(\mathrm{II})} \in \mathbb{R}$ we have* *(i)* $\limsup_{N\to\infty} \Pi_{\theta_N}^N\Big(\frac{\mathcal{U}_N^{(II)}-b_N^{(II)}}{a_N^{(II)}} \le w\Big) \le U_2^{(II)}(w)$ *for all* $w \in \mathbb{R}$,

(ii) $\liminf_{N\to\infty} \Pi_{\theta_N}^N\Big(\frac{\mathcal{U}_N^{(II)}-b_N^{(II)}}{a_N^{(II)}} \le w\Big) \ge U_1^{(II)}(w)$ *for all* $w \in [w_0^{(II)}, \infty)$. *Then (II.2) holds true with* $b_{\mathcal{U}^{(II)}}(\Pi_{\theta_N}^N) = b_N^{(II)}$.

Further suppose that $w_0^{(II)}$ *is such that for some* $w_1^{(II)} > w_0^{(II)}$ *satisfying* $0 < U_1^{(II)}(w_0^{(II)}) < U_2^{(II)}(w_1^{(II)}) < 1$ *it holds for* N *sufficiently large*

(iii)
$$\sup_{P\times Q\in\mathbb{P}_0} P^{m(N)} \times Q^{n(N)}\left(\frac{\mathcal{U}_N^{(II)} - b_N^{(II)}}{a_N^{(II)}} > w_1^{(II)}\right) < \sup_{P\times Q\in\mathbb{P}_0} P^{m(N)} \times Q^{n(N)}\left(\frac{\mathcal{U}_N^{(II)} - b_N^{(II)}}{a_N^{(II)}} > w_0^{(II)}\right).$$

Finally assume that (II.1) holds for $\mathcal{U}_N^{(II)}$ *and for the above* $\{b_N^{(II)}\}$ *and* $\{\gamma_N\}$, $\{\lambda_N\}$ *appearing in (II.1) it holds* $[b_N^{(II)}]^2/\lambda_N \to 0$ *and* $[b_N^{(II)}]^2/\gamma_N \to \infty$. *Then, (2.11) is satisfied with*

$$\alpha_N = \sup_{P\times Q\in\mathbb{P}_0} P^{m(N)} \times Q^{n(N)}\left(\frac{\mathcal{U}_N^{(II)} - b_N^{(II)}}{a_N^{(II)}} > w_1^{(II)}\right) \tag{B.1}$$

while the asymptotic power of the pertaining test based on $\mathcal{U}^{(II)}$ *lies in the interval* $\Big[1 - U_2^{(II)}(w_1^{(II)}), 1 - U_1^{(II)}(w_0^{(II)})\Big]$.

For completeness we state also a simple analogue of Lemma B.1 which may be useful for checking (I.2) for $\mathcal{U}_N^{(I)}$. Its proof is quite similar to that of Lemma B.1, so we omit it.

Lemma B.2. *Let* $\{P_{\vartheta_N} \times Q_{\vartheta_N}\}$ *be arbitrary sequence of alternatives for which* $\vartheta_N \to 0$ *and* $N\vartheta_N^\rho \to \infty$, $\rho \in [1, 2]$. *Set* $b_N^{(I)} = \sqrt{mn/N}\, b_{\mathcal{U}^{(I)}}(P_{\vartheta_N} \times Q_{\vartheta_N})$, *where the positive function* $b_{\mathcal{U}^{(I)}}(P_\vartheta \times Q_\vartheta)$ *is defined for all* $\vartheta \in (0, 1)$. *Suppose that, for each* $\{\vartheta_N\}$ *as above, there exist cumulative distribution functions* $U_1^{(I)}$ *and* $U_2^{(I)}$ *and a positive sequence* $\{a_N^{(I)}\}$ *such that* $b_N^{(I)}/a_N^{(I)} \to \infty$ *and for some* $w_0^{(I)} \in \mathbb{R}$ *we have*

(i) $\limsup_{N\to\infty} \Pi_{\vartheta_N}^N\Big(\frac{\mathcal{U}_N^{(I)}-b_N^{(I)}}{a_N^{(I)}} \le w\Big) \le U_2^{(I)}(w)$ *for all* $w \in \mathbb{R}$,

(ii) $\liminf_{N\to\infty} \Pi_{\vartheta_N}^N\Big(\frac{\mathcal{U}_N^{(I)}-b_N^{(I)}}{a_N^{(I)}} \le w\Big) \ge U_1^{(I)}(w)$ *for all* $w \in [w_0^{(I)}, \infty)$.

Then (I.2) is satisfied with the above $b_{\mathcal{U}^{(I)}}(\cdot)$. *Distribution functions* $U_1^{(I)}$ *and* $U_2^{(I)}$, *the sequence* $\{a_N^{(I)}\}$ *as well as* $w_0^{(I)}$ *may be different for each sequence* $\{\vartheta_N\}$.

Note that if for all N sufficiently large $\sup_{P\times Q\in\mathbb{P}_0} P^{m(N)} \times Q^{n(N)}\big(\mathcal{U}_N^{(II)} - b_N^{(II)} < w a_N^{(II)}\big)$ is strictly increasing in w then (iii) holds true for every $w_1^{(II)} > w_0^{(II)}$.

Proof of Lemma B.1. First we shall check that indeed the conditions (i) and (ii) yield (2.10) with $b_{\mathcal{U}^{(II)}}(\Pi_{\theta_N}^N) = b_N^{(II)}$. We have

$$\Pi_{\theta_N}^N\left(\Big|\frac{\mathcal{U}_N^{(II)}}{b_N^{(II)}} - 1\Big| \ge \epsilon\right) = \Pi_{\theta_N}^N\left(\Big|\frac{\mathcal{U}_N^{(II)} - b_N^{(II)}}{a_N^{(II)}}\Big| \ge \frac{b_N^{(II)}}{a_N^{(II)}}\epsilon\right). \tag{B.2}$$

Since $b_N^{(\mathrm{II})}/a_N^{(\mathrm{II})} \to \infty$ we can take $w^* > w_0^{(\mathrm{II})}$, and N enough large, to majorize limsup of (B.2) by $1 - U_1^{(\mathrm{II})}(w^*) + U_2^{(\mathrm{II})}(-w^*)$. Since w^* can be arbitrary large the bound is arbitrary small.

Now we shall check that $\{\alpha_N\}$ given in (B.1) fulfills the requirements needed to calculate the intermediate efficiency *via* Theorem 2.6.

By the definition of $u_{\alpha_N N}^{(\mathrm{II})}$ and (iii) it follows

$$b_N^{(\mathrm{II})} + a_N^{(\mathrm{II})} w_0^{(\mathrm{II})} \le u_{\alpha_N N}^{(\mathrm{II})} \le b_N^{(\mathrm{II})} + a_N^{(\mathrm{II})} w_1^{(\mathrm{II})}. \tag{B.3}$$

Due to the assumptions on the sequences $\{a_N^{(\mathrm{II})}\}$, $\{b_N^{(\mathrm{II})}\}$, $\{\gamma_N\}$ and $\{\lambda_N\}$ we have

$$(b_N^{(\mathrm{II})} + a_N^{(\mathrm{II})} w_1^{(\mathrm{II})})^2 = [b_N^{(\mathrm{II})}]^2 \left(1 + \frac{a_N^{(\mathrm{II})}}{b_N^{(\mathrm{II})}} w_1^{(\mathrm{II})}\right)^2 \to \infty$$

and

$$\frac{(b_N^{(\mathrm{II})} + a_N^{(\mathrm{II})} w_1^{(\mathrm{II})})^2}{\lambda_N} = \frac{[b_N^{(\mathrm{II})}]^2}{\lambda_N} \left(1 + \frac{a_N^{(\mathrm{II})}}{b_N^{(\mathrm{II})}} w_1^{(\mathrm{II})}\right)^2 \to 0,$$

$$\frac{(b_N^{(\mathrm{II})} + a_N^{(\mathrm{II})} w_1^{(\mathrm{II})})^2}{\gamma_N} = \frac{[b_N^{(\mathrm{II})}]^2}{\gamma_N} \left(1 + \frac{a_N^{(\mathrm{II})}}{b_N^{(\mathrm{II})}} w_1^{(\mathrm{II})}\right)^2 \to \infty.$$

Hence, for $w_N^2 = (b_N^{(\mathrm{II})} + a_N^{(\mathrm{II})} w_1^{(\mathrm{II})})^2/N$ the condition (II.1) can be applied and yields

$$-\frac{1}{(b_N^{(\mathrm{II})} + a_N^{(\mathrm{II})} w_1^{(\mathrm{II})})^2} \log \sup_{P \times Q \in \mathbb{P}_0} P^{m(N)} \times Q^{n(N)} (\mathcal{U}_N^{(\mathrm{II})} \ge b_N^{(\mathrm{II})} + a_N^{(\mathrm{II})} w_1^{(\mathrm{II})}) \to c_{\mathcal{U}^{(\mathrm{II})}}.$$

By (B.3) and the definition of α_N, this implies

$$-\frac{\log \alpha_N}{(b_N^{(\mathrm{II})} + a_N^{(\mathrm{II})} w_1^{(\mathrm{II})})^2} \to c_{\mathcal{U}^{(\mathrm{II})}}.$$

Similar argument works for $w_1^{(\mathrm{II})}$ replaced by $w_0^{(\mathrm{II})}$. This shows that $\{\alpha_N\}$ satisfies the condition (2.11). Moreover, again by (B.3),

$$\Pi_{\theta_N}^N \left(\frac{\mathcal{U}_N^{(\mathrm{II})} - b_N^{(\mathrm{II})}}{a_N^{(\mathrm{II})}} \ge w_1^{(\mathrm{II})} \right) \le \Pi_{\theta_N}^N (\mathcal{U}_N^{(\mathrm{II})} \ge u_{\alpha_N N}^{(\mathrm{II})}) \le \Pi_{\theta_N}^N \left(\frac{\mathcal{U}_N^{(\mathrm{II})} - b_N^{(\mathrm{II})}}{a_N^{(\mathrm{II})}} \ge w_0^{(\mathrm{II})} \right).$$

Taking appropriate limits of both sides we infer that the above chosen sequence $\{\alpha_N\}$, in addition to satisfy (2.11), belongs to $\mathbb{L}^*$, as

$$0 < 1 - U_2^{(\mathrm{II})}(w_1^{(\mathrm{II})}) \le \liminf_{N\to\infty} \Pi_{\theta_N}^N \left(\mathcal{U}_N^{(\mathrm{II})} \ge u_{\alpha_N N}^{(\mathrm{II})}\right) \le \limsup_{N\to\infty} \Pi_{\theta_N}^N \left(\mathcal{U}_N^{(\mathrm{II})} \ge u_{\alpha_N N}^{(\mathrm{II})}\right)$$

$$\le 1 - U_1^{(\mathrm{II})}(w_0^{(\mathrm{II})}) < 1.$$

□

Appendix C. Proof of Theorem 3.1

The argument follows the idea developed in Inglot and Ledwina [20] and exploits the Komlós–Major–Tusnády inequality for the uniform empirical process. Therefore, consider two probability spaces, two independent sequences $\{B_m^{'}\}$ and $\{B_n^{''}\}$ of Brownian bridges defined on them, and two independent sequences of uniform empirical processes $\{e_m^{'}\}$ and $\{e_n^{''}\}$, defined on the same spaces, such that for all m, n and $w \in \mathbb{R}$

$$Pr\left(\sup_{t\in[0,1]} |e_m^{'}(t) - B_m^{'}(t)| \geq \frac{w + C\log m}{\sqrt{m}}\right) \leq L\exp\{-lw\},$$

$$Pr\left(\sup_{t\in[0,1]} |e_n^{''}(t) - B_n^{''}(t)| \geq \frac{w + C\log n}{\sqrt{n}}\right) \leq L\exp\{-lw\},$$

where C, L and l are absolute positive constants. Here and in what follows Pr denotes a probability measure pertaining to the underlying probability space. On the other hand,

$$\sqrt{\frac{mn}{N}}\Big\{\hat{G}_n(z) - \hat{F}_m(z)\Big\} \stackrel{D}{=} \sqrt{\frac{m}{N}}e_n^{''}(J_1(z)) - \sqrt{\frac{n}{N}}e_m^{'}(J_1(z)),$$

where $\stackrel{D}{=}$ denotes the equality in distribution while $B_N^0 \stackrel{D}{=} \sqrt{\frac{m}{N}}B_n^{''} - \sqrt{\frac{n}{N}}B_m^{'}$ is a Brownian bridge. Hence, by the above and the property $Pr\big(\sup_{t\in[0,1]} B_N^0(t) \geq w\big) = \exp\{-2w^2\},\ w \in \mathbb{R}$, we get

$$P_0^N(\mathcal{V}_N \geq w_N\sqrt{N}) = Pr\left(\sup_{t\in[0,1]}\left\{\sqrt{\frac{m}{N}}e_n^{''}(t) - \sqrt{\frac{n}{N}}e_m^{'}(t)\right\} \geq w_N\sqrt{N}\right)$$

$$\leq Pr\left(\sup_{t\in[0,1]} B_N^0(t) \geq (1-\sqrt{w_N})w_N\sqrt{N}\right)$$

$$+Pr\left(\sup_{t\in[0,1]} |e_m^{'}(t) - B_m^{'}(t)| \geq \sqrt{\frac{N}{n}}\frac{\sqrt{w_N}}{2}w_N\sqrt{N}\right) + Pr\left(\sup_{t\in[0,1]} |e_n^{''}(t) - B_n^{''}(t)| \geq \sqrt{\frac{N}{m}}\frac{\sqrt{w_N}}{2}w_N\sqrt{N}\right)$$

$$\leq (1+o(1))\exp\big\{-2(1-\sqrt{w_N})^2 w_N^2 N\big\}.$$

Analogously we obtain $P_0^N(\mathcal{V}_N \geq w_N\sqrt{N}) \geq (1+o(1))\exp\big\{-2(1+\sqrt{w_N})^2 w_N^2 N\big\}$. □

Appendix D. Proof of Theorem 3.2 and verification of (3.10)

Since we like to apply some results of Inglot [19] therefore we have to adjust our statistics to pertaining ones considered in that paper. First of all note that the results of that paper apply, as well, to rank statistics with the score function depending on N.

Next observe that, by (i) and (ii), it holds everywhere

$$\max_{1\leq j\leq\Delta(N)}\left|\sum_{i=1}^N c_{Ni}\,\ell_j\left(\frac{R_i - 0.5}{N}\right) - \sum_{i=1}^N c_{Ni}\,\ell_j\left(\frac{R_i}{N}\right)\right| \leq$$

$$\max_{1\leq j\leq\Delta(N)} \sqrt{\frac{N}{mn}}\frac{1}{\sqrt{\pi_{jN}(1-\pi_{jN})}} \leq \frac{\Delta(N)+1}{\sqrt{\Delta(N)}}\sqrt{\frac{N}{mn}} = O\left(\sqrt{\frac{\Delta(N)}{N}}\right) = o(1). \quad \text{(D.1)}$$

So, we can abandon the correction for continuity in $\mathcal{L}_j$.

Finally, we construct appropriate continuous approximation of the score functions $\ell_j,\ j = 1, 2, .., \Delta(N)$. For this purpose some auxiliary notation are introduced.

For a fixed $\tau \in (0, 1)$ set

$$l(t;\tau) = -\sqrt{\frac{1-\tau}{\tau}}\,\mathbf{1}(0 \le t < \tau) + \sqrt{\frac{\tau}{1-\tau}}\,\mathbf{1}(\tau \le t \le 1).$$

Given $\epsilon \in (0, \tau(1-\tau))$ we shall modify l on the interval $\mathbb{I}_\epsilon(\tau) = [\tau(1-\epsilon), \tau(1-\epsilon)+\epsilon]$ containing the jump point τ. To this end introduce the function $r(t;\tau,\epsilon)$ which is 0 outside $\mathbb{I}_\epsilon(\tau)$,

$$r(t;\tau,\epsilon) = \sqrt{\frac{1-\tau}{\tau}} + \frac{1}{\epsilon}\sqrt{\frac{1-\tau}{\tau^3}}(t-\tau) \quad \text{if} \quad \tau(1-\epsilon) \le t < \tau$$

and

$$r(t;\tau,\epsilon) = -\sqrt{\frac{\tau}{1-\tau}} + \frac{1}{\epsilon}\sqrt{\frac{\tau}{(1-\tau)^3}}(t-\tau) \ \text{if} \ \ \tau \le t \le \tau(1-\epsilon)+\epsilon.$$

Then define

$$\bar{l}(t;\tau,\epsilon) = \sqrt{\frac{3}{3-2\epsilon}}\Big[l(t;\tau) + r(t;\tau,\epsilon)\Big]$$

and note that $\bar{l}(t;\tau,\epsilon)$ is piecewise linear, absolutely continuous, and satisfies $\int_0^1 \bar{l}(t;\tau,\epsilon)dt = 0$, and $\int_0^1 \bar{l}^2(t;\tau,\epsilon)dt = 1$. Moreover, on the interval $\mathbb{I}_\epsilon(\tau)$ it holds that $|l(t;\tau) - \bar{l}(t;\tau,\epsilon)| \le 1/\sqrt{\tau(1-\tau)}$ while outside this interval $|l(t;\tau) - \bar{l}(t;\tau,\epsilon)| \le \epsilon/\sqrt{\tau(1-\tau)}$. For $\epsilon < 1/N$ there is at most one point R_i/N in the interval $\mathbb{I}_\epsilon(\tau)$. Hence

$$\left|\sum_{i=1}^N c_{Ni}\, l\left(\frac{R_i}{N};\tau\right) - \sum_{i=1}^N c_{Ni}\, \bar{l}\left(\frac{R_i}{N};\tau,\epsilon\right)\right| \le \Big(\epsilon + \lceil N\epsilon\rceil\Big)\frac{1}{\sqrt{\tau(1-\tau)}}\sqrt{\frac{N}{mn}}. \tag{D.2}$$

Take now $\tau = \pi_{jN},\ \epsilon = \epsilon_N = 1/(2N)$ and define

$$\bar{\mathcal{L}}_j = \bar{\mathcal{L}}_{jN} = \sum_{i=1}^N c_{Ni}\,\bar{l}\left(\frac{R_i}{N};\pi_{jN},\epsilon_N\right) \qquad \text{and} \qquad \bar{\mathcal{T}}_N = \max_{1\le j\le \Delta(N)}\{-\bar{\mathcal{L}}_j\}.$$

Then, by (D.1) and (D.2), for all N large enough we have everywhere

$$\big|\mathcal{L}_j - \bar{\mathcal{L}}_j\big| \le \frac{2}{\sqrt{\eta(1-\eta)}}\sqrt{\Delta(N)/N} \quad \text{ and } \quad \big|\mathcal{T}_N - \bar{\mathcal{T}}_N\big| \le \frac{2}{\sqrt{\eta(1-\eta)}}\sqrt{\Delta(N)/N}. \tag{D.3}$$

For each of the rank statistic $\bar{\mathcal{L}}_j,\ j = 1, \dots, \Delta(N)$, we shall apply Theorem 3.4 of Inglot [19]. Note that in our situation we need to insert there $\Psi(1)$ in place of $\Psi(d(N))$, where $\Psi(1) = \int_0^1 \big|\frac{\partial}{\partial t}\bar{l}(t;\pi_{jN},\epsilon_N)\big|dt$; *cf.* (3.13) ibidem. We have $\Psi(1) = \sqrt{3}/\sqrt{(3-2\epsilon_N)\pi_{jN}(1-\pi_{jN})}$. Moreover, λ_N appearing in that theorem equals 1 in our application. The above yields

$$P_0^N\big(\big|\bar{\mathcal{L}}_j\big| \ge w_N\sqrt{N}\big) = \exp\Big\{-\frac{1}{2}Nw_N^2 + O(Nw_N^{2+\upsilon/2}) + O(\log Nw_N^2)\Big\} \tag{D.4}$$

uniformly in j. This implies that

$$P_0^N\left(\bar{\mathcal{T}}_N \geq w_N\sqrt{N}\right) = \exp\Big\{-\frac{1}{2}Nw_N^2 + O(Nw_N^{2+\upsilon/2}) + O(\log Nw_N^2) + O(\log\Delta(N))\Big\}. \tag{D.5}$$

In view of (D.3) and (i), (D.5) yields

$$P_0^N\left(\mathcal{T}_N \geq w_N\sqrt{N}\right) = \exp\Big\{-\frac{1}{2}Nw_N^2 + O(Nw_N^{2+\upsilon/2}) + O(\log Nw_N^2) + O(\log\Delta(N))\Big\}. \tag{D.6}$$

Since $\Delta(N) = o(N)$, then by (iii), $O(\log\Delta(N)) + O(\log Nw_N^2) = o(Nw_N^2)$. Hence (3.12) follows. □

Verification of (3.10)

We argue similarly as in the proof of Lemma A.1 in Ledwina and Wyłupek [36].

Put

$$\widetilde{\mathcal{L}}_j = -\sqrt{N/mn}\,\mathcal{L}_j.$$

Let $Z_1,\dots,Z_N$ denote the pooled sample $X_1,\dots,X_m,Y_1,\dots,Y_n$ and let $Z_{(r)}$ stand for the r-th order statistic of the pooled sample. For any $j = 1,\dots,\Delta(N)$ we have

$$\widetilde{\mathcal{L}}_j = -\int_{Z_{(1)}}^{\infty} l_j\left(\hat{J}_N(x) - \frac{1}{2N}\right)d(\hat{G}_n(x) - \hat{F}_m(x)) = -\int_{1/N}^{1} l_j\left(t - \frac{1}{2N}\right)d(\hat{G}_n - \hat{F}_m)\circ\hat{J}_N^{-1}(t).$$

Applying to the last expression the integration by parts formula, *cf.* (1) in Shorack [47], page 115, we get

$$\widetilde{\mathcal{L}}_j = (\hat{G}_n - \hat{F}_m)\circ\hat{J}_N^{-1}\left(\frac{1}{N}\right)l_j\left(\frac{1}{2N}\right) + \int_{1/N}^{1}(\hat{G}_n - \hat{F}_m)\circ\hat{J}_N^{-1}(t)\,d\,l_j\left(t - \frac{1}{2N}\right)$$

$$= (\hat{G}_n - \hat{F}_m)\circ\hat{J}_N^{-1}\left(\frac{1}{N}\right)l_j\left(\frac{1}{2N}\right) + \frac{1}{\sqrt{\pi_{jN}(1-\pi_{jN})}}(\hat{G}_n - \hat{F}_m)\circ\hat{J}_N^{-1}\left(\pi_{jN} + \frac{1}{2N}\right).$$

Set

$$\widetilde{W}_j = \frac{1}{\sqrt{\pi_{jN}(1-\pi_{jN})}}(\widehat{G}_n - \widehat{F}_m)\circ\widehat{J}_N^{-1}(\pi_{jN}).$$

Then we have

$$|\widetilde{\mathcal{L}}_j - \widetilde{W}_j| \leq \frac{1}{\sqrt{\pi_{jN}(1-\pi_{jN})}}\left[\left|(\hat{G}_n - \hat{F}_m)\circ\hat{J}_N^{-1}\left(\frac{1}{N}\right)\max\{\pi_{jN}, 1-\pi_{jN}\}\right|\right.$$

$$\left. + \left|(\hat{G}_n - \hat{F}_m)\circ\hat{J}_N^{-1}\left(\pi_{jN} + \frac{1}{2N}\right) - (\hat{G}_n - \hat{F}_m)\circ\hat{J}_N^{-1}(\pi_{jN})\right|\right]. \tag{D.7}$$

By the definition, $\widehat{J}_N^{-1}(t)$ equals $Z_{(\lceil N\,t\rceil)}$, $t \in (0,1)$. So, the first term in (D.7) is majorized by

$$\frac{1}{\sqrt{\pi_{jN}(1-\pi_{jN})}}\max\left\{\frac{1}{m}, \frac{1}{n}\right\}. \tag{D.8}$$

When $\lceil N\pi_{jN}+0.5\rceil=\lceil N\pi_{jN}\rceil$ then the second term in (D.7) equals 0. When $\lceil N\pi_{jN}+0.5\rceil=\lceil N\pi_{jN}\rceil+1$ then the second term in (D.7) is also majorized by (D.8) Hence

$$|\widetilde{\mathcal{L}}_j-\widetilde{\mathcal{W}}_j|\le\frac{2}{\sqrt{\pi_{jN}(1-\pi_{jN})}}\max\left\{\frac{1}{m},\frac{1}{n}\right\}.$$

Since

$$\mathcal{T}_N=\sqrt{\frac{mn}{N}}\max_{1\le j\le\Delta(N)}\widetilde{L}_j\quad\text{and}\quad\mathcal{W}_N=\sqrt{\frac{mn}{N}}\max_{1\le j\le\Delta(N)}\widetilde{W}_j,$$

by the triangle inequality, we get

$$|\mathcal{T}_N-\mathcal{W}_N|\le\sqrt{\frac{mn}{N}}\max_{1\le j\le\Delta(N)}|\widetilde{\mathcal{L}}_j-\widetilde{\mathcal{W}}_j|$$
$$\le\frac{1}{\sqrt{N}}\max\left\{\sqrt{\frac{1-\eta_N}{\eta_N}},\sqrt{\frac{\eta_N}{1-\eta_N}}\right\}\times\frac{2}{\min\{\sqrt{\pi_{1N}(1-\pi_{1N})},\sqrt{\pi_{\Delta(N)N}(1-\pi_{\Delta(N)N})}\}}.$$

This, after elementary argument, yields (3.10). □

Appendix E. Proof of Theorem 3.3

We have $\sqrt{m}\{\hat{F}_m-F_{1N}\}\overset{D}{=}e_m^{(1)}\circ F_{1N},\ \ \sqrt{n}\{\hat{G}_n-G_{1N}\}\overset{D}{=}e_n^{(2)}\circ G_{1N}$, where $e_m^{(1)}$ and $e_n^{(2)}$ are independent uniform empirical processes defined on an appropriate probability space. In particular, one can use the KMT constructions applied in the proof of Theorem 3.1. Hence

$$\sqrt{\frac{mn}{N}}\{\hat{G}_n-\hat{F}_m\}\overset{D}{=}\sqrt{\frac{m}{N}}e_n^{(2)}\circ G_{1N}-\sqrt{\frac{n}{N}}e_m^{(1)}\circ F_{1N}+\sqrt{\frac{mn}{N}}\vartheta_N(G_1-F_1)\tag{E.1}$$

and

$$\Pi^N_{\vartheta_N}(\mathcal{V}_N-b_{\mathcal{V}}(\Pi^N_{\vartheta_N})\le w)\le Pr\left(\sqrt{\frac{m}{N}}e_n^{(2)}\circ G_{1N}(z_0)-\sqrt{\frac{n}{N}}e_m^{(1)}\circ F_{1N}(z_0)\le w\right).\tag{E.2}$$

Since $G_{1N}(z_0)\to J_1(z_0)$ and $F_{1N}(z_0)\to J_1(z_0)$, therefore the random variable on the right hand side of (E.2) has asymptotic $N(0,\sqrt{J_1(z_0)[1-J_1(z_0)]})$ law. This justifies the form of V_2.

On the other hand, by (E.1) we infer that

$$\Pi^N_{\vartheta_N}\left(\mathcal{V}_N-b_{\mathcal{V}}(\Pi^N_{\vartheta_N})\le w\right)\ge\Pi^N_{\vartheta_N}\left(\sqrt{\frac{mn}{N}}\sup_{z\in\mathbb{R}}\left\{\hat{G}_n(z)-\hat{F}_m(z)-\vartheta_N[G_1(z)-F_1(z)]\right\}\le w\right)=$$

$$Pr\left(\sup_{z\in\mathbb{R}}\left\{\sqrt{\frac{m}{N}}e_n^{(2)}\circ G_{1N}(z)-\sqrt{\frac{n}{N}}e_m^{(1)}\circ F_{1N}(z)\right\}\le w\right).$$

Since $e_m^{(1)}\circ F_{1N}\Rightarrow B^{(1)}$ and $e_n^{(2)}\circ G_{1N}\Rightarrow B^{(2)}$, where $B^{(1)}$ and $B^{(2)}$ are independent Brownian bridges while $\Rightarrow$ denotes weak convergence, the form of V_1 follows. □

Appendix F. Proof of Theorem 3.4

F.1. Preliminaries. We shall prove (3.16) and (3.17) for the statistic $\mathcal{W}_N$. Since (ii) implies that $\Delta(N) = o(\sqrt{N}/\log N)$, therefore (3.10) justifies such approach.

Set

$$\kappa_N = \sqrt{\eta_N(1-\eta_N)} \quad \text{and} \quad \nu_N = \eta_N - \eta$$

and introduce two auxiliary processes on $\mathbb{R}$

$$\zeta_N(z) = \sqrt{N}\kappa_N[\hat{G}_n(z) - G_{1N}(z)] - \sqrt{N}\kappa_N[\hat{F}_m(z) - F_{1N}(z)],$$

$$\xi_N(z) = \sqrt{N}(1-\eta_N)[\hat{G}_n(z) - G_{1N}(z)] + \sqrt{N}\eta_N[\hat{F}_m(z) - F_{1N}(z)].$$

For $\hat{J}_N(z) = \eta_N\hat{F}_m(z) + (1-\eta_N)\hat{G}_n(z)$ put

$$\hat{z}_{jN} = \hat{J}_N^{-1}(\pi_{jN}).$$

Additionally, set $V_N(z) = \sqrt{N}\kappa_N\{\hat{G}_n(z) - \hat{F}_m(z)\}$. With these notation

$$\hat{J}_N(z) = \frac{1}{\sqrt{N}}\xi_N(z) + J_1(z) - \theta_N\nu_N\bar{A}(J_1(z)) \tag{F.1}$$

while

$$V_N(z) = \zeta_N(z) + \sqrt{N}\theta_N\kappa_N\bar{A}(J_1(z)) \quad \text{and} \quad \mathcal{W}_N = \max_{1\le j\le\Delta(N)} \frac{V_N(\hat{z}_{jN})}{\sqrt{\pi_{jN}(1-\pi_{jN})}}. \tag{F.2}$$

Now, let us reperametrize F_{1N} and G_{1N} in (3.2) to a classical form in the two-sample scheme, which we shall exploit below. For $t \in (0,1)$ set

$$\bar{A}(t) = \bar{A}(t;\eta) = (G_1 - F_1)\circ J_1^{-1}(t), \quad \bar{A}^+(t) = \bar{A}^+(t;\eta) = \max\{\bar{A}(t), 0\}. \tag{F.3}$$

With the above notation, (3.2) can be written as

$$F_{1N} = J_1 - \vartheta_N(1-\eta)\bar{A}\circ J_1, \quad G_{1N} = J_1 + \vartheta_N\eta\bar{A}\circ J_1. \tag{F.4}$$

By Remark 2.2, $\bar{a}(t) = \bar{A}'(t)$ exists almost everywhere (with respect to the Lebesgue measure) and it holds that

$$-\eta^{-1} \le \bar{a}(t) \le (1-\eta)^{-1} \quad \text{and} \quad \int_0^1 \bar{a}(t)dt = 0;$$

cf. Behnen and Neuhaus [5], for example. Note also that

$$\frac{dF_{1N}}{dJ_1}\circ J_1^{-1}(t) = 1 - \vartheta_N(1-\eta)\bar{a}(t) \quad \text{and} \quad \frac{dG_{1N}}{dJ_1}\circ J_1^{-1}(t) = 1 + \vartheta_N\eta\bar{a}(t). \tag{F.5}$$

In consequence, for each $\eta \in (0,1)$ we have $\bar{A}(0;\eta) = \bar{A}(1;\eta) = 0$ and

$$\lim_{t\to 0+} A^*(t;\eta) = \lim_{t\to 1-} A^*(t;\eta) = 0, \quad \text{where } A^*(t;\eta) = \frac{\bar{A}(t;\eta)}{\sqrt{t(1-t)}}. \tag{F.6}$$

By (F.6) there exists $\delta \in (0,1/2)$ such that

$$\max_{j:\pi_{jN}\notin[2\delta,1-2\delta]} \frac{\bar{A}(\pi_{jN})}{\sqrt{\pi_{jN}(1-\pi_{jN})}} \le \frac{1}{2}\mu_0, \quad \text{where} \quad \mu_0 = \max_{j:\pi_{jN}\in[2\delta,1-2\delta]} \frac{\bar{A}(\pi_{jN})}{\sqrt{\pi_{jN}(1-\pi_{jN})}}. \tag{F.7}$$

To increase readability of the proof of (3.16) and (3.17) we formulate now some partial results, which we shall justify at Sections F.2–F.6.

For δ defined *via* (F.7) set

$$z_1 = J_1^{-1}(\delta) \quad \text{and} \quad z_2 = J_1^{-1}(1-\delta).$$

Recall that $\Pi^N_{\theta_N} = P^{m(N)}_{\theta_N} \times Q^{n(N)}_{\theta_N}$, where P_{θ_N} and Q_{θ_N} are defined *via* (F.4) with $\theta_N \to 0$ and $N\theta_N^2 \to \infty$. In the succeeding lemmas we specify sufficient conditions on $\{\theta_N\}$ for them to hold. Throughout C is an absolute constant, not necessarily the same in all places.

Lemma F.1.
(a) *If $\theta_N \to 0$ and (ii) holds then*

$$\lim_{N\to\infty} \Pi^N_{\theta_N}\left(\sup_{z\in[z_1,z_2]} \frac{|\zeta_N(z)|}{\sqrt{\hat{J}_N(z)[1-\hat{J}_N(z)]}} \le w \right) = Pr\left(\sup_{t\in[\delta,1-\delta]} \frac{|B(t)|}{\sqrt{t(1-t)}} \le w \right), \ w \in \mathbb{R}_+, \tag{F.8}$$

where B is a Brownian bridge.
(b) *Assume that (i), (ii) and (iii) of Theorem 3.4 hold. Set*

$$\mathbb{E}_{0N} = \left\{ \sup_{z\in\mathbb{R}} |\hat{J}_N(z) - J_1(z)| \le \sqrt{\frac{\log N}{N}} \right\}, \quad \mathbb{E}_{1N} = \left\{ \sup_{z\in\mathbb{R}} |\hat{J}_N(z) - J_1(z)| \le \frac{C}{\theta_N \Delta(N)\sqrt{N}} \right\},$$

and

$$\mathbb{E}_{2N} = \left\{ \max_{1\le j\le\Delta(N)} \frac{|\bar{A}(\hat{J}_N(\hat{z}_{jN})) - \bar{A}(J_1(\hat{z}_{jN}))|}{\sqrt{\pi_{jN}(1-\pi_{jN})}} \le \frac{1}{\theta_N\sqrt{N\Delta(N)}} \right\}.$$

Then

$$\lim_{N\to\infty} \Pi^N_{\theta_N}\left(\mathbb{E}_{0N}\right) = \lim_{N\to\infty} \Pi^N_{\theta_N}\left(\mathbb{E}_{1N}\right) = \lim_{N\to\infty} \Pi^N_{\theta_N}\left(\mathbb{E}_{2N}\right) = 1. \tag{F.9}$$

Moreover, the following useful bounds take place. On $\mathbb{E}_{0N}$ we have

$$|J_1(\hat{z}_{jN}) - \pi_{jN}| \le |J_1(\hat{z}_{jN}) - J_N(\hat{z}_{jN})| + |J_N(\hat{z}_{jN}) - \pi_{jN}| \le \sqrt{\frac{\log N}{N}} + \frac{1}{N} \le 2\sqrt{\frac{\log N}{N}} \tag{F.10}$$

while on $\mathbb{E}_{1N}$

$$|J_1(\hat{z}_{jN}) - \pi_{jN}| \le \frac{C}{\theta_N \Delta(N)\sqrt{N}}. \tag{F.11}$$

Further introduce

$$l_N = J_1^{-1}\left(\frac{\log N}{\sqrt{N}}\right) \quad \text{and} \quad u_N = J_1^{-1}\left(1 - \frac{\log N}{\sqrt{N}}\right)$$

and note that for N large enough it holds $l_N \le u_N$.

Lemma F.2. *Suppose that* $N\theta_N^2/\log^2 N \to \infty$, $\theta_N\sqrt{N}\nu_N = O(1)$, *and* $w_N \asymp \theta_N\sqrt{N}$. *Then*

$$\lim_{N\to\infty} \Pi_{\theta_N}^N\left(\sup_{z\in[l_N,u_N]} \frac{|\zeta_N(z)|}{\sqrt{\hat{J}_N(z)[1-\hat{J}_N(z)]}} \le w_N\right) = 1. \tag{F.12}$$

Lemma F.3. *Under (i), (ii) and (iii) of Theorem 3.4, for* $\mathbb{E}_{3N}$ *given by*

$$\mathbb{E}_{3N} = \left\{\mathcal{W}_N \le \max_{j:\hat{z}_{jN}\in[z_1,z_2]} \frac{V_N(\hat{z}_{jN})}{\sqrt{\pi_{jN}(1-\pi_{jN})}}\right\}$$

it holds that

$$\lim_{N\to\infty} \Pi_{\theta_N}^N\left(\mathbb{E}_{3N}\right) = \lim_{N\to\infty} \Pi_{\theta_N}^N\left(\max_{j:\hat{z}_{jN}\notin[z_1,z_2]} \frac{V_N(\hat{z}_{jN})}{\sqrt{\pi_{jN}(1-\pi_{jN})}} \le \max_{j:\hat{z}_{jN}\in[z_1,z_2]} \frac{V_N(\hat{z}_{jN})}{\sqrt{\pi_{jN}(1-\pi_{jN})}}\right) = 1. \tag{F.13}$$

By the above, to prove (3.16) and (3.17) it is enough to consider

$$\Pi_{\theta_N}^N\left(\left\{\mathcal{W}_N - b_{\mathcal{T}}(\Pi_{\theta_N}^N) \le w\right\} \cap \bigcap_{j=0}^{3} \mathbb{E}_{jN}\right).$$

F.2. Proof of (3.16). Let $j_0 = j_0(N)$ be any index j such that

$$\max_{1\le j\le\Delta(N)} \frac{\bar{A}(\pi_{jN})}{\sqrt{\pi_{jN}(1-\pi_{jN})}} = \frac{\bar{A}(\pi_{j_0N})}{\sqrt{\pi_{j_0N}(1-\pi_{j_0N})}}$$

By (F.7), without loss of generality we can assume that j_0 is such that for each N it holds that $\pi_{j_0N} \in [2\delta, 1-2\delta]$. With this notation

$$b_{\mathcal{T}}(\Pi_{\theta_N}^N) = \sqrt{N}\theta_N\kappa_N \frac{\bar{A}(\pi_{j_0N})}{\sqrt{\pi_{j_0N}(1-\pi_{j_0N})}}.$$

By (F.6), (F.11) and (i), on the set $\mathbb{E}_{1N}$

$$\mathcal{W}_N - b_{\mathcal{T}}(\Pi^N_{\theta_N}) \geq \frac{\zeta_N(\hat{z}_{j_0N}) + \sqrt{N}\theta_N\kappa_N[\bar{A}(J_1(\hat{z}_{j_0N})) - \bar{A}(\pi_{j_0N})]}{\sqrt{\pi_{j_0N}(1-\pi_{j_0N})}} = \frac{\zeta_N(\hat{z}_{j_0N})}{\sqrt{\pi_{j_0N}(1-\pi_{j_0N})}} + o(1).$$

Therefore, to conclude the proof of (3.16) it is enough to show that

$$\frac{\zeta_N(\hat{z}_{j_0N})}{\sqrt{\pi_{j_0N}(1-\pi_{j_0N})}} \xrightarrow{D} N(0,1) \quad \text{as} \quad N \to \infty. \tag{F.14}$$

The main difficulty in proving (F.14) lies in that $j_0 = j_0(N)$ may be not unique and changes with N. Therefore, we proceed as follows. When N is growing then, by (i), the partition is getting more dense. Hence, the set of accumulation points of the sequence $\{\pi_{j_0N}\}$ is nonempty and is contained in $[2\delta, 1-2\delta]$. Therefore, it is enough to prove (F.14) for any concentration point and pertaining subsequence of $\{\pi_{j_0N}\}$ converging to it. Set t_0 to be any concentration point of the sequence and denote by $\{\pi_{j_0'N'}\}$, $j_0' = j_0(N')$, a subsequence converging to t_0. By the definition of $\hat{z}_{jN}$ and (F.10), on $\mathbb{E}_{0N}$ it holds

$$|J_1(\hat{z}_{j_0'N'}) - t_0| \leq |J_1(\hat{z}_{j_0'N'}) - \hat{J}_{N'}(\hat{z}_{j_0'N'})| + |\hat{J}_{N'}(\hat{z}_{j_0'N'}) - \pi_{j_0'N'}| + |\pi_{j_0'N'} - t_0| \leq 2\sqrt{\frac{\log N}{N}} + |\pi_{j_0'N'} - t_0|.$$

and yields

$$J_1(\hat{z}_{j_0'N'}) \xrightarrow{\Pi^{N'}_{\theta_{N'}}} t_0.$$

This and the continuity of J_1, imply that subsequence $\{\hat{z}_{j_0'N'}\}$ converges in $\Pi^{N'}_{\theta_{N'}}$ to $J_1^{-1}(t_0)$. Hence, weak convergence of the process $\zeta_N(z)/\sqrt{\hat{J}_N(z)[1-\hat{J}_N(z)]}$, to the process $B(J_1(z))/\sqrt{J_1(z)[1-J_1(z)]}$, $z \in [z_1, z_2]$, *cf.* the proof of Lemma F.1, implies that, under $\Pi^{N'}_{\theta_{N'}}$,

$$\frac{\zeta_{N'}(\hat{z}_{j_0'N'})}{\sqrt{\pi_{j_0'N'}(1-\pi_{j_0'N'})}} \xrightarrow{D} \frac{B(t_0)}{\sqrt{t_0(1-t_0)}} \quad \text{as} \quad N' \to \infty. \tag{F.15}$$

This shows that for any convergent subsequence $\{\pi_{j_0'N'}\}$ of the sequence $\{\pi_{j_0N}\}$ the sequence of random variables in (F.15) converges to the same limiting N(0,1) law. This proves (3.16). □

F.3. Proof of (3.17). Recall that we can restrict attention to $\bigcap_{j=0}^3 \mathbb{E}_{jN}$. In particular, on $\mathbb{E}_{1N} \cap \mathbb{E}_{3N}$, by (F.6), (F.11) and Lipschitz condition for $\bar{A}$, we have

$$\mathcal{W}_N - b_{\mathcal{T}}(\Pi^N_{\theta_N}) \leq \max_{j:\hat{z}_{jN}\in[z_1,z_2]} \frac{V_N(\hat{z}_{jN}) - \sqrt{N}\theta_N\kappa_N\bar{A}(\pi_{jN})}{\sqrt{\pi_{jN}(1-\pi_{jN})}} = \max_{j:\hat{z}_{jN}\in[z_1,z_2]} \Bigg\{ \frac{\zeta_N(\hat{z}_{jN})}{\sqrt{\pi_{jN}(1-\pi_{jN})}} +$$

$$\frac{\sqrt{N}\theta_N\kappa_N[\bar{A}(J_1(\hat{z}_{jN})) - \bar{A}(\pi_{jN})]}{\sqrt{\pi_{jN}(1-\pi_{jN})}} \Bigg\} \leq \max_{j:\hat{z}_{jN}\in[z_1,z_2]} \left\{ \frac{|\zeta_N(\hat{z}_{jN})|}{\sqrt{\pi_{jN}(1-\pi_{jN})}} \right\} + \frac{C}{\sqrt{\Delta(N)}}. \tag{F.16}$$

Now observe that the property $|\hat{J}_N(\hat{z}_{jN})-\pi_{jN}|<1/N$ implies that $|\hat{J}_N(\hat{z}_{jN})[1-\hat{J}_N(\hat{z}_{jN})]-\pi_{jN}(1-\pi_{jN})|\le 1/N$. Hence, by (i), for N large enough

$$1-2\frac{\Delta(N)}{N}\le\sqrt{\frac{\hat{J}_N(\hat{z}_{jN})[1-\hat{J}_N(\hat{z}_{jN})]}{\pi_{jN}(1-\pi_{jN})}}\le 1+2\frac{\Delta(N)}{N}. \tag{F.17}$$

Hence, the right hand side of (F.16) is majorized by

$$[1+o(1)]\sup_{z\in[z_1,z_2]}\frac{|\zeta_N(z)|}{\sqrt{\hat{J}_N(z)[1-\hat{J}_N(z)]}}+o(1).$$

By (F.8) of Lemma F.1 the proof is concluded. □

F.4. Proof of Lemma F.1.

(a) As in the proof of Theorem 3.3, an application of strong approximation technique implies that $\zeta_N\Rightarrow B^{(1)}\circ J_1$ and $\xi_N\Rightarrow B^{(2)}\circ J_1$, where $B^{(1)}$ and $B^{(2)}$ are independent Brownian bridges. Moreover, (F.1) implies that $\hat{J}_N\overset{\Pi^N_{\theta_N}}{\longrightarrow}J_1$. Hence (F.8) follows.

(b) By (F.1) it holds

$$\Pi^N_{\theta_N}(\mathbb{E}^c_{0N})\le\Pi^N_{\theta_N}\left(\sup_{z\in\mathbb{R}}|\xi_N(z)-\theta_N\nu_N\sqrt{N}\bar{A}(J_1(z))|>\sqrt{\log N}\right)$$

$$\le\Pi^N_{\theta_N}\left(\sup_{z\in\mathbb{R}}|\xi_N(z)|\ge\sqrt{\log N}-C\theta_N\nu_N\sqrt{N}\right)$$

and, by the weak convergence of ξ_N and the assumption (iii), $\lim_{N\to\infty}\Pi^N_{\theta_N}(\mathbb{E}^c_{0N})=0$. Analogously,

$$\Pi^N_{\theta_N}(\mathbb{E}^c_{1N})\le\Pi^N_{\theta_N}\left(\sup_{z\in\mathbb{R}}|\xi_N(z)-\theta_N\sqrt{N}\nu_N\bar{A}(J_1(z))|>\frac{C}{\theta_N\Delta(N)}\right)$$

$$\le\Pi^N_{\theta_N}\left(\sup_{z\in\mathbb{R}}|\xi_N(z)|\ge\frac{C}{\theta_N\Delta(N)}+O(1)\right)$$

and we infer that $\Pi^N_{\theta_N}(\mathbb{E}^c_{1N})\to 0$. Moreover, since $\bar{A}$ is Lipschitz one, then, with some C, we have

$$\Pi^N_{\theta_N}(\mathbb{E}^c_{2N})\le\Pi^N_{\theta_N}\left(\frac{\sup_{z\in\mathbb{R}}|\hat{J}_N(z)-J_1(z)|}{\min_{1\le j\le\Delta(N)}\sqrt{\pi_{jN}(1-\pi_{jN})}}>\frac{C}{\theta_N\sqrt{N\Delta(N)}}\right).$$

Thus, by (i), it holds $\Pi^N_{\theta_N}(\mathbb{E}^c_{2N})\le\Pi^N_{\theta_N}(\mathbb{E}^c_{1N})$, with some appropriate C in $\mathbb{E}_{1N}$. Hence, the proof of (F.9) is completed. □

F.5. Proof of Lemma F.2. On $\mathbb{E}_{0N}$, given in Lemma F.1, for $z \in (0,1)$ it holds $|\hat{J}_N(z)[1-\hat{J}_N(z)] - J_1(z)[1-J_1(z)]| \le \sqrt{\log N/N}$. Hence, by the definition of l_N and u_N, for $z \in [l_N, u_N]$

$$\Big|\frac{\hat{J}_N(z)[1-\hat{J}_N(z)]}{J_1(z)[1-J_1(z)]} - 1\Big| \le \frac{4}{\sqrt{\log N}}.$$

For N large enough this implies

$$\Pi^N_{\theta_N}\left(\sup_{z\in[l_N,u_N]} \frac{|\zeta_N(z)|}{\sqrt{\hat{J}_N(z)[1-\hat{J}_N(z)]}} \ge w_N\right)$$

$$\le \Pi^N_{\theta_N}\left(\sup_{z\in[l_N,u_N]} \frac{|\zeta_N(z)|}{\sqrt{J_1(z)[1-J_1(z)]}} \ge w_N\big(1-\frac{4}{\sqrt{\log N}}\big)\right) + o(1). \tag{F.18}$$

As in the proof of Theorem 3.1, consider now the uniform empirical process $e^{'}_m$ and related Brownian Bridge $B^{'}_m$ and independent on them $e^{''}_n$ and $B^{''}_n$ such that the KMT inequalities hold for them. Under $\Pi^N_{\theta_N}$, $\sqrt{m}[\hat{F}_m - F_{1N}] \overset{D}{=} e^{'}_m(F_{1N})$, $\sqrt{n}[\hat{G}_n - G_{1N}] \overset{D}{=} e^{''}_n(G_{1N})$ and $\zeta_N \overset{D}{=} \tilde{\zeta}_N = \sqrt{\frac{m}{N}}e^{''}_n(G_{1N}) - \sqrt{\frac{n}{N}}e^{'}_m(F_{1N})$. Set $B_N = \sqrt{\frac{m}{N}}B^{''}_m - \sqrt{\frac{n}{N}}B^{'}_n$. Then B_N is a Brownian bridge. Therefore, we can majorize the first component of (F.18) as follows

$$Pr\left(\sup_{z\in[l_N,u_N]} \frac{|B_N(J_1(z))|}{\sqrt{J_1(z)[1-J_1(z)]}} \ge w_N\left(\frac{1}{2} - \frac{4}{\sqrt{\log N}}\right)\right)$$

$$+ Pr\left(\sup_{z\in[l_N,u_N]} \frac{|\tilde{\zeta}_N(z) - B_N(J_1(z))|}{\sqrt{J_1(z)[1-J_1(z)]}} \ge \frac{w_N}{2}\right). \tag{F.19}$$

Due to the definition of l_N, u_N, the assumptions $w_N \asymp \theta_N\sqrt{N}, \theta_N\sqrt{N}/\log N \to \infty$, Darling and Erdős result, *cf.* Lemma 4.4.1 in Csörgő *et al.* [8], implies that the first component of (F.19) tends to 0.

Using again the form of l_N and u_N, the second component of (F.19) for large N is majorized by

$$Pr\Big(\sup_{z\in\mathbb{R}} |\tilde{\zeta}_N(z) - B_N(J_1(z))| \ge w^*_N\Big), \tag{F.20}$$

where $w^*_N = w_N\sqrt{\log N}/(4N^{1/4})$. The structure of $\tilde{\zeta}_N$, given above, allows to majorize (F.20) as follows

$$\begin{aligned}
&Pr\Big(\sup_{z\in\mathbb{R}} |e^{'}_m(F_{1N}(z)) - B^{'}_m(J_1(z))| \ge \kappa_N w^*_N\Big) + Pr\Big(\sup_{z\in\mathbb{R}} |e^{''}_n(G_{1N}(z)) - B^{''}_n(J_1(z))| \ge \kappa_N w^*_N\Big) \\
&\le Pr\Big(\sup_{t\in(0,1)} |e^{'}_m(t) - B^{'}_m(t)| \ge \kappa_N w^*_N/2\Big) + Pr\Big(\sup_{t\in(0,1)} |e^{''}_n(t) - B^{''}_n(t)| \ge \kappa_N w^*_N/2\Big) \\
&+2Pr\Big(\sup_{0\le t\le 1-C\theta_N}\ \sup_{0\le h\le C\theta_N} |B^{'}_m(t+h) - B^{'}_m(t)| \ge \kappa_N w^*_N/2\Big).
\end{aligned} \tag{F.21}$$

Since $\eta_N \to \eta, w_N \asymp \theta_N\sqrt{N}$ and $\theta_N\sqrt{N} \to \infty$ an application of the KMT inequality to the two first components of (F.21) shows that these terms are negligible. The last term of (F.21) requires standard analysis of increments of the Brownian bridge. Applying for this purpose Lemma A of Inglot [19] with $h = C\theta_N, y = y_N =$

$\kappa_N w_N^*/(2\sqrt{C\theta_N})$ and $\delta = 1/4$ finishes the proof, as $y \to \infty$ faster than $\log N$. □

F.6. Proof of Lemma F.3. Recall that $v(t) = \sqrt{t(1-t)}$. Note that

$$\mathbb{E}_{3N}^c \subset \left\{ \max_{j:\hat{z}_{jN}\notin[z_1,z_2]} \frac{V_N(\hat{z}_{jN})}{v(\pi_{jN})} > \max_{j:\hat{z}_{jN}\in[z_1,z_2]} \frac{V_N(\hat{z}_{jN})}{v(\pi_{jN})} \right\}.$$

Throughout we restrict attention to $\mathbb{E}_{0N}$. By (F.10), for sufficiently large N, $\{j : z_1 \le \hat{z}_{jN} \le z_2\} = \{j : \delta \le J_1(\hat{z}_{jN}) \le 1-\delta\} \supset \{j : 2\delta \le \pi_{jN} \le 1-2\delta\}$. Hence $\{j : \hat{z}_{jN} \notin [z_1, z_1]\} \subset \{j : \pi_{jN} \notin [2\delta, 1-2\delta]\}$ and

$$\max_{j:\hat{z}_{jN}\notin[z_1,z_2]} \frac{V_N(\hat{z}_{jN})}{v(\pi_{jN})} \le \max_{j:\hat{z}_{jN}\notin[z_1,z_2]} \frac{V_N(\hat{z}_{jN}) - \sqrt{N}\theta_N\kappa_N\bar{A}(\pi_{jN})}{v(\pi_{jN})} + \max_{j:\pi_{jN}\notin[2\delta,1-2\delta]} \frac{\sqrt{N}\theta_N\kappa_N\bar{A}(\pi_{jN})}{v(\pi_{jN})}. \tag{F.22}$$

Using (F.6), (F.7), (F.10) and (iii) we conclude

$$\max_{j:\hat{z}_{jN}\notin[z_1,z_2]} \frac{V_N(\hat{z}_{jN})}{v(\pi_{jN})} \le \max_{j:\hat{z}_{jN}\notin[z_1,z_2]} \frac{|V_N(\hat{z}_{jN}) - \sqrt{N}\theta_N\kappa_N\bar{A}(\pi_{jN})|}{v(\pi_{jN})} + \frac{1}{2}\sqrt{N}\theta_N\kappa_N\mu_0$$

$$\le \max_{j:\hat{z}_{jN}\notin[z_1,z_2]} \frac{|\zeta_N(\hat{z}_{jN})|}{v(\pi_{jN})} + \rho_N^{(2)}, \tag{F.23}$$

where $\rho_N^{(2)} = \sqrt{N}\theta_N\kappa_N(\mu_0/2 + C\sqrt{\Delta(N)\log N/N}) \asymp \sqrt{N}\theta_N$.

Analogously,

$$\max_{j:\hat{z}_{jN}\in[z_1,z_2]} \frac{V_N(\hat{z}_{jN})}{v(\pi_{jN})} \ge \max_{j:\pi_{jN}\in[2\delta,1-2\delta]} \frac{\sqrt{N}\theta_N\kappa_N\bar{A}(\pi_{jN})}{v(\pi_{jN})} - \max_{j:\hat{z}_{jN}\in[z_1,z_2]} \frac{|V_N(\hat{z}_{jN}) - \sqrt{N}\theta_N\kappa_N\bar{A}(\pi_{jN})|}{v(\pi_{jN})}$$

$$\ge \rho_N^{(1)} - \max_{j:\hat{z}_{jN}\in[z_1,z_2]} \frac{|\zeta_N(\hat{z}_{jN})|}{v(\pi_{jN})}, \tag{F.24}$$

where $\rho_N^{(1)} = \sqrt{N}\theta_N\kappa_N(\mu_0 - C\sqrt{\Delta(N)\log N/N}) \asymp \sqrt{N}\theta_N$.

The above implies that

$$\mathbb{E}_{0N} \cap \mathbb{E}_{3N}^c \subset \left\{ \max_{1\le j\le\Delta(N)} \frac{|\zeta_N(\hat{z}_{jN})|}{\sqrt{\pi_{jN}(1-\pi_{jN})}} > \frac{1}{2}\left[\rho_N^{(1)} - \rho_N^{(2)}\right] \right\}.$$

Now, observe that, by (i), (ii) and (iii), it follows that

$$\mathbb{E}_{0N} \subset \bigcap_{j=1}^{\Delta(N)} \left\{ \hat{z}_{jN} \in [l_N, u_N] \right\} \quad \text{and} \quad \lim_{N\to\infty} \Pi_{\theta_N}^N \left(\mathbb{E}_{0N} \cap \bigcap_{j=1}^{\Delta(N)} \left\{ \hat{z}_{jN} \in [l_N, u_N] \right\} \right) = 1. \tag{F.25}$$

Indeed, by (ii), $\Delta(N) = o(\sqrt{N}/\log N)$. Hence, for N large enough, we have $1/[\Delta(N)+1] > [3\log N]/\sqrt{N}$. Therefore, (F.10) and (i) imply that

$$\log N/\sqrt{N} \le J_1(\hat{z}_{1N}) \le J_1(\hat{z}_{\Delta(N)N}) \le 1 - \log N/\sqrt{N}.$$

By (F.25) we infer

$$\Pi^N_{\theta_N}\left(\mathbb{E}_{0N} \cap \mathbb{E}^c_{3N}\right) \le \Pi^N_{\theta_N}\left(\sup_{z\in[l_N,u_N]} \frac{|\zeta_N(\hat{z}_{jN})|}{\sqrt{\hat{J}_N(z)[1-\hat{J}_N(z)]}} \ge \rho_N[1+o(1)]\right),$$

where $\rho_N = [\rho_N^{(1)} - \rho_N^{(2)}]/2$ and $\rho_N \asymp \theta_N\sqrt{N}$. An application of (B.8) finishes the proof. □

Appendix G. Proof of Theorem 3.5

We shall argue that Theorems 3.1–3.4 imply, *via* Lemmas B.1 and B.2, that the regularity assumptions (I.1), (I.2), (II.1) and (II.2) hold true. Besides, (II.1) holds with such $\{\gamma_N\}$ and $\{\lambda_N\}$ that (2.11) is satisfied and Theorem 2.6 works.

For $\mathcal{U}_N^{(\mathrm{I})} = \mathcal{V}_N$ the situation is easy. Theorem 3.1 implies (I.1) while Theorem 3.3 along with Lemma B.2 yield (I.2).

To verify (II.1) for $\mathcal{U}_N^{(\mathrm{II})} = \mathcal{T}_N$ it is enough to indicate sequences $\{\gamma_N\}$ and $\{\lambda_N\}$ such that Theorem 3.2 yields (II.1). Observe that $\gamma_N = \log N$ and $\lambda_N = N/\Delta(N)^{2/(1-\nu)}$ are adequate. Indeed, $Nw_N^2/\lambda_N = [w_N^{1-\nu}\Delta(N)]^{2/(1-\nu)} = o(1)$ and Theorem 3.2 applies with $v = \nu$. Similarly, $Nw_N^2/\gamma_N \to \infty$. Hence (II.1) is proved.

Assumptions of Theorem 3.5 are stronger than that of Theorem 3.4. Therefore, by Theorem 3.4, (i) and (ii) of Lemma B.1 hold true with $U_2^{(\mathrm{II})}(w) = \Phi(w)$, $U_1^{(\mathrm{II})}(w) = T_1(w)$. Pertaining sequence $\{b_N^{(\mathrm{II})}\} = \{b_{\mathcal{T}}(\Pi^N_{\theta_N})\}$ is of the order $\theta_N\sqrt{N}$. The distribution function T_1 has the property $K(\sqrt{\delta[1-\delta]}w) \le T_1(w) \le 2\Phi(w) - 1$, where $K(w) = Pr(\sup_{0<t<1}|B(t)| \le w)$. This implies that $T_1(0) = 0$ and, by Tsirel'son [49], $T_1(w)$ is absolutely continuous on $[0,\infty)$. This allows for choosing $w_0^{(\mathrm{II})}$ arbitrarily close to 0 and proves (II.2).

Since the distribution of $\mathcal{T}_N$ is discrete one and its atoms depend on N, therefore (iii) of Lemma B.1 deserves some comment. Recall that $\mathcal{T}_N = \max_{1\le j\le\Delta(N)}\{-\mathcal{L}_j\}$; *cf.* (3.6). Due to stochastic monotonicity of $\mathcal{T}_N$ one can restrict attention to the case $F = G$. Then the distribution of the vector of ranks is uniform. By (3.4) and (3.5), for each $j = 1,\dots,\Delta(N)$ it holds

$$\mathcal{L}_j = \sqrt{\frac{N}{mn}}\frac{1}{\sqrt{\pi_{jN}(1-\pi_{jN})}}\left[\sum_{i=1}^m \frac{n}{N}\mathbf{1}_{[0,\pi_{jN})}\left(\frac{R_i - 0.5}{N}\right) - \sum_{i=m+1}^N \frac{m}{N}\mathbf{1}_{[0,\pi_{jN})}\left(\frac{R_i - 0.5}{N}\right)\right].$$

Note that the value of $\mathcal{L}_j$ depends only on the number of $(R_i - 0.5)/N,\ i = 1,\dots,m$, falling into $[0,\pi_{jN})$. Hence, if this number increases by 1 then the first sum in $\mathcal{L}_j$ increases by n/N while the second one decreases by m/N. In consequence, the value of $\mathcal{L}_j$ increases by

$$\delta_{jN} = \sqrt{\frac{N}{mn}}\frac{1}{\sqrt{\pi_{jN}(1-\pi_{jN})}} \le \delta_N = 4\sqrt{\frac{\Delta(N)}{N}},\ \ j = 1,\dots,\Delta(N).$$

Most sparse are locations of atoms of $\mathcal{L}_1$ and $\mathcal{L}_{\Delta(N)}$. The minimal value of $\mathcal{L}_1$ is attained when in the interval $[0, \pi_{1N}]$ ranks of the observations from the first sample are absent. This minimal value, say L_1, satisfies

$$-L_1 = \lfloor N\pi_{1N} + 0.5 \rfloor \frac{m}{N} \sqrt{\frac{N}{mn}} \frac{1}{\sqrt{\pi_{1N}(1-\pi_{1N})}} \asymp \sqrt{\frac{N\pi_{1N}}{1-\pi_{1N}}} \geq \sqrt{\frac{N}{\Delta(N)}}.$$

Since $b_N^{(\mathrm{II})} \asymp \theta_N \sqrt{N}$ the assumption (ii)' yields $|L_1/b_N^{(\mathrm{II})}| \to \infty$.

Similar argument applies to $\mathcal{T}_N$ and yields that the atoms of the distribution of this statistic are located at points with distance not exceeding the above defined δ_N. Hence, in any interval of a fixed length, lying right to the point $b_N^{(\mathrm{II})} = b_{\mathcal{T}}(\Pi_{\theta_N}^N)$, there is at least one value of $\mathcal{T}_N$ and (iii) of Lemma B.1 holds.

Finally, since $b_N^{(\mathrm{II})} \asymp \theta_N \sqrt{N}$, the assumption (ii)' implies that $[b_N^{(\mathrm{II})}]^2/\lambda_N \to 0$ and $[b_N^{(\mathrm{II})}]^2/\gamma_N \to \infty$ as $N \to \infty$. Therefore, by Lemma B.1, (2.11) holds true with α_N given in (B.1). Since $\max_{1\le j\le \Delta(N)}\{\pi_{jN} - \pi_{j-1N}\} \to 0$, (2.12) holds, as well, and proves (3.18). ☐

Acknowledgements. The authors would like to thank to two referees for carefully reading the paper and for suggestions, which led to an improved version of the paper. The paper was partially written when B. Ćmiel was on leave from AGH University of Science and Technology and was granted by postdoc position at the Institute of Mathematics of the Polish Academy of Sciences. Moreover, the work of B. Ćmiel was partially supported by the Faculty of Applied Mathematics AGH UST dean grant for PhD students and young researchers within subsidy of Ministry of Science and Higher Education.

References

[1] R.R. Bahadur, Rates of convergence of estimates and test statistics. *Ann. Math. Statist.* **38** (1967) 303–324.
[2] G.F. Barrett and S.G. Donald, Consistent tests for stochastic dominance. *Econometrica* **71** (2003) 71–104.
[3] K. Behnen, A characterization of certain rank-order tests with bounds for the asymptotic relative efficiency. *Ann. Math. Statist.* **43** (1972) 1839–1851.
[4] K. Behnen, Nichtparametrische Statistik: Zweistichproben Rangtests. *Z. Angew. Math. Mech.* **61** (1981) T203–T212.
[5] K. Behnen and G. Neuhaus, Galton's test as a linear rank test with estimated scores and its local asymptotic efficiency. *Ann. Statist.* **11** (1983) 588–599.
[6] K. Behnen and G. Neuhaus, Rank Tests with Estimated Scores and Their Application. Teubner, Stuttgart (1989).
[7] A.A. Borovkov and A.A. Mogulskii, Large deviations and statistical invariance principle. *Theory Probab. Appl.* **37** (1993) 7–13.
[8] M. Csörgő, S. Csörgő, L. Horváth and D.M. Mason, Weighted empirical and quantile processes. *Ann. Probab.* **14** (1986) 31–85.
[9] G.R. Ducharme and T. Ledwina, Efficient and adaptive nonparametric test for the two-sample problem. *Ann. Statist.* **31** (2003) 2036–2058.
[10] M.S. Ermakov, Large deviations for empirical probability measures and statistical tests. *J. Math. Sci.* **81** (1996) 2379–2393.
[11] M.S. Ermakov, On asymptotically efficient statistical inference for moderate deviation probabilities. *Theory Probab. Appl.* **48** (2004) 622–641.
[12] J. Fan, Test of significance based on wavelet thresholding and Neyman's truncation. *J. Am. Statist. Assoc.* **96** (1996) 647–688.
[13] J.-D. Fermanian, D. Radulović and M. Wegkamp, Weak convergence of empirical copula process. *Bernoulli* **10** (2004) 847–860.
[14] P. Groeneboom and G.R. Shorack, Large deviations of goodness of fit statistics and linear combinations of order statistics. *Ann. Probab.* **9** (1981) 971–987.
[15] M.S. Handcock and M. Morris, Relative Distribution Methods in the Social Sciences. Springer, New York (1999).
[16] T. Inglot, Generalized intermediate efficiency of goodness of fit tests. *Math. Methods Statist.* **8** (1999) 487–509.
[17] T. Inglot, On large deviation theorem for data-driven Neyman's statistic. *Statist. Probab. Lett.* **47** (2000) 411–419.
[18] T. Inglot, Intermediate efficiency by shifting alternatives and evaluation of power. *J. Statist. Plan. Inference* **140** (2010) 3263–3281.
[19] T. Inglot, Asymptotic behaviour of linear rank statistics for the two-sample problem. *Probab. Math. Statist.* **32** (2012) 93–116.
[20] T. Inglot and T. Ledwina, On probabilities of excessive deviations for Kolmogorov-Smirnov, Cramér-von Mises and chi-square statistics. *Ann. Statist.* **18** (1990) 1491–1495.
[21] T. Inglot and T. Ledwina, Moderately large deviations and expansions of large deviations for some functionals of weighted empirical process. *Ann. Probab.* **21** (1993) 1691–1705.

- [22] T. Inglot and T. Ledwina, Asymptotic optimality of data driven Neyman's tests for uniformity. *Ann. Statist.* **24** (1996) 1982–2019.
- [23] T. Inglot and T. Ledwina, Intermediate approach to comparison of some goodness-of-fit tests. *Ann. Inst. Statist. Math.* **53** (2001) 810–834.
- [24] T. Inglot and T. Ledwina, Intermediate efficiency of some max-type statistics. *J. Statist. Plan. Inference* **136** (2006) 2918–2935.
- [25] L. Jager and J.A. Wellner, On the "Poisson boundaries" of the family of weighted Kolmogorov statistics, in Festschrift for Herman Rubin, edited by A. DasGupta. IMS, Beachwood, OH (2004) 319–331.
- [26] L. Jager and J.A. Wellner, Goodness-of-fit tests via phi-divergences. *Ann. Statist.* **35** (2007) 2018–2053.
- [27] G. Kalish and P.W. Mikulski, The asymptotic behavior of the Smirnov test compared to standard "optimal procedures". *Ann. Math. Statist.* **42** (1971) 1742–1747.
- [28] W.C.M. Kallenberg, Intermediate efficiency, theory and examples. *Ann. Statist.* **11** (1983) 1401–1420.
- [29] W.C.M. Kallenberg, Efficiency, intermediate or Kallenberg, in Encyclopedia of Statistical Sciences, edited by S. Kotz *et al.*, Wiley, New York (1999) 1876–1881.
- [30] W.C.M. Kallenberg and A.J. Koning, On Wieand's theorem. *Statist. Probab. Lett.* **25** (1995) 113–132.
- [31] Y. Kitamura, Asymptotic optimality of empirical likelihood for testing moment restrictions. *Econometrica* **69** (2001) 1661–1672.
- [32] S. Klonner, The first-order stochastic dominance ordering of the Singh-Maddala distribution. *Econ. Lett.* **69** (2000) 123–128.
- [33] A.J. Koning, Approximation of stochastic integrals with applications to goodness-of-fit. *Ann. Statist.* **20** (1992) 428–454.
- [34] T. Ledwina, On the limiting Pitman efficiency of some rank tests of independence. *J. Multivar. Anal.* **20** (1986) 265–271.
- [35] T. Ledwina and G. Wyłupek, Nonparametric tests for first order stochastic dominance. *TEST* **21** (2012) 730–756.
- [36] T. Ledwina and G. Wyłupek, Two-sample test against one-sided alternative. *Scand. J. Statist.* **39** (2012) 358–381.
- [37] T. Ledwina and G. Wyłupek, Tests for first-order stochastic dominance. Preprint IM PAN (2013) 746.
- [38] D.M. Mason and R.L. Eubank, Moderate deviations and intermediate efficiency for lack-of-fit tests. *Statist. Risk Model.* **29** (2012) 175–187.
- [39] S.M. Mirakhmedov, Asymptotic intermediate efficiency of the chi-square and likelihood ratio goodness of fit tests. Preprint arXiv:1610.04135 (2016).
- [40] G. Neuhaus, H_0-contiguity in nonparametric testing problems and sample Pitman efficiency. *Ann. Statist.* **10** (1982) 575–582.
- [41] G. Neuhaus, Local asymptotics for linear rank statistics with estimated score functions. *Ann. Statist.* **15** (1987) 491–512.
- [42] Y. Nikitin, Asymptotic Efficiency of Nonparametric Tests. Cambridge University Press, Cambridge (1995).
- [43] E.J.G. Pitman, Notes on Nonparametric Statistical Inference. Institute of Statistics, University of North Carolina, Chapel Hill (1948).
- [44] E.J.G. Pitman, Some Basic Theory for Statistical Inference. Wiley, NY (1979).
- [45] F. Schmid and M. Trede, Testing for first order stochastic dominance: a new distribution-free test. *Statistician* **45** (1996) 371–380.
- [46] R.J. Serfling, Approximation Theorems of Mathematical Statistics. Wiley, New York (1980).
- [47] G.R. Shorack, Probability for Statisticians. Springer, New York (2000).
- [48] B.K. Sinha and H.S. Wieand, Bounds on the efficiencies of four commonly used nonparametric tests of location. *Sankhyā, Ser. B* **39** (1977) 121–129.
- [49] V.S. Tsirel'son, The density of the distribution of the maximum of a Gaussian process. *Theory Probab. Appl.* **20** (1975) 847–856.
- [50] H.S. Wieand, A condition under which the Pitman and Bahadur approaches to efficiency coincide. *Ann. Statist.* **4** (1976) 1003–1011.
- [51] C.S. Yu, Pitman efficiencies of Kolmogorov-Smirnov tests. *Ann. Math. Statist.* **42** (1971) 1595–1605.
- [52] S. Zacks, Pitman efficiency, in Encyclopedia of Statistical Sciences. Edited by S. Kotz *et al.* Wiley, New York (2006) 6136–6140.