TI 2005-122/1
Tinbergen Institute Discussion Paper

Comparing Distributions:

The Harmonic Mass Index

Jeroen Hinloopen[a]

Charles van Marrewijk[b]

[a] Universiteit van Amsterdam, and Tinbergen Institute;
[b] Erasmus Universiteit Rotterdam, and Tinbergen Institute.

Comparing distributions: The Harmonic Mass index

Jeroen Hinloopen[*]

University of Amsterdam (and *Tinbergen Institute*)

Charles van Marrewijk[†‡]

Erasmus University Rotterdam (and *Tinbergen Institute*)

December 2005

Abstract

The information contained in PP-plots is transformed into a single number. The resulting Harmonic Mass (HM) index is distribution free and its sample counterpart is shown to be consistent. For a wide class of CDFs the exact analytical expression of the distribution of the sample HM index is derived, assuming the two underlying samples to be drawn from the same distribution. The robustness of the concomitant test statistic is assessed, and four different methods are discussed for applying the HM test in case of asymmetric samples.

Key words: Distribution, PP-plot, test statistic, critical percentile values, robustness.

JEL Classification: C12, C14

[*]Correspondence: University of Amsterdam, Faculty of Economics and Econometrics, Amsterdam School of Economics, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands; `J.Hinloopen@uva.nl`; www.fee.uva.nl/io/jhinloopen.

[†]Correspondence: Erasmus University Rotterdam, Rotterdam School of Economics, Department of Economics, P.O. Box 1728, 3000 DR Rotterdam, The Netherlands; `vanmarrewijk@few.eur.nl`; www.charlesvanmarrewijk.nl.

1   Introduction

Research abounds involving comparisons of distributions. In all areas of economics – empirical, experimental, and theoretical – these comparisons are used to substantiate conclusions. Do macroeconomic variables assume different values across countries? Do different experiment treatments yield different outcomes? Do predictions differ between model simulations based on varying sets of parameter values? And so on.

In many cases the theoretical distribution of the underlying economic phenomenon is not known and the comparison of (sample) distributions boils down to the comparison of (sample) statistics. Next to the traditional consideration of sample moments and extreme values, the most-frequently used test in applied research involves some version of the Mann-Whitney-Wilcoxon (Mww) U statistic.[1] As this elegant test yields the probability that one variable is less than or equal to another, it is not suited for testing whether two (sample) distributions differ as such; rejecting the hypothesis that neither of two random variables exceeds the other does not imply acceptance of the hypothesis that they are drawn from the same distribution. This paper introduces a statistic for testing just that.

Other comparisons typically involve graphical displays, most notably quantile-quantile (QQ) plots and percentile-percentile (PP) plots. The former is the scatter plot of quantiles of two distributions for all entries of their joint support while the latter displays for all these values the scatter plot of the two distributions' percentiles. If distributions differ in scale and location only QQ-plots consist of straight lines with slope one, which by some is considered to be a desirable feature (Wilk and Gnanadesikan, 1968). Yet, they become degenerate in case of non-overlapping supports. Relatedly, QQ-plots are highly sensitive to observations in either distributions' tails. While this readily allows for the identification of outliers it blurs the pattern that is present in the majority of the data. Also, if one support encompasses the other QQ-plots can be straight lines (due to the scale-location-invariance property) even though the probability mass is distributed quite differently for the two underlying distributions.

Although PP-plots do not carry the scale-location-invariance property they have several advantages. First, comparisons between distributions are always available in compact space as PP-plots are scale-invariant. Second, they are little obscured by outlying observations because they are much more sensitive to differences in the centre of the mass of the underlying distrib-

---

[1] An alternative to the Mww U statistic is the Kuiper or Kolmogorov-Smirnov test. As this statistic is based on the maximum point-wise deviation between distributions it is highly sensitive to outliers thus reducing its power substantially.

utions than to differences in their tails. This property in particular makes them well-suited for establishing whether two samples are drawn from the same distribution as the implicit focus is on the pattern exhibited by the majority of the data. Third, and most importantly, PP-plots contain all necessary information to make scale-invariant comparisons between distributions (Holmgren, 1995).

The informative features of graphical displays notwithstanding, they do not provide information from which statistical inferences can be drawn.[2] Indeed, in this paper the information of a PP-plot is translated into a single number between 0 and 1. Because a PP-plot is the diagonal in case the two underlying distributions are identical, an obvious measure for characterizing PP-plots is the area between the diagonal and the plot. It is this area, multiplied by two, that is considered here. As it reflects the extent to which the probability mass of the two underlying distributions is in harmony, it is labelled the Harmonic Mass (HM) index.

Theoretically the HM index is readily defined, as is the PP-plot. For practical use the sample HM index is to be considered, which specifies the index value for any two discrete samples. The sample HM index is shown to be consistent. Absent within-sample and between-sample ties, simple rules are specified to compute the index value without approximation error.

Next, the probability density function (pdf) of the sample HM index is derived under the assumption that the two underlying samples are drawn from the same distribution. This readily allows for hypotheses testing. As this pdf involves burdensome computations, critical percentiles are provided of the related cumulative density function (CDF) for samples including up to 350 entries. For larger samples an accurate approximation rule is derived.

To some extent the use of the HM test is ruled by its inherent Type-I and Type-II error. Outliers are shown to have a mild influence on the HM test; simulated critical percentiles are hardly affected by data contamination of up to five percent of all observations, while those at 97.5 and 99 are little affected by a corruption of 10 percent of the data. This is comfortably in line with Hampel *et al.* (1986) who find that 10 percent is the maximum fraction of pollution in routine data. The acceptance rate of the HM test in case the two underlying distributions differ, depends on the specific distributions that are compared. Simulations indicate that the HM test is quite able to discern the log-normal, uniform, and exponential distribution in case these distributions are parameterized such that they mimic each other.

Finally, an assessment of the distribution dynamics of the Balassa index

---

[2]Another use of PP-plots is to compare different plots. Holmgren (1995) provides semiparametric and nonparametric tests for these comparisons.
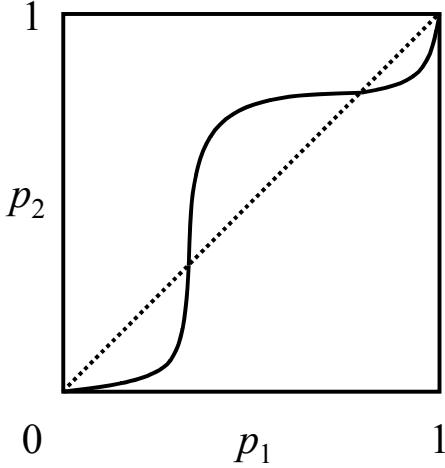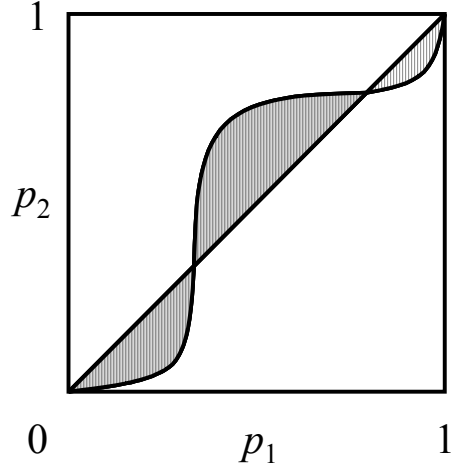
Figure 1: Theoretical PP-plot (Panel a) and concomitant Harmonic Mass index (Panel b).

for revealed comparative advantage in the United Kingdom (UK) illustrates the use of the HM test. In this case it reveals a particular era to be relatively volatile, an era that is overlooked by a traditional Markov analysis.

2   The Harmonic Mass index

Consider the following set of functions: $\Xi_1 = \{ F | \forall x, h \in \mathbb{R}: \lim_{x \to -\infty} F(x) = 0, \lim_{x \to \infty} F(x) = 1, \lim_{h \downarrow 0} F(x + h) = F(x)$, and $a < b \implies F(a) \leq F(b) \}$. Let F and G be CDFs belonging to Ξ_1. The concomitant PP-plot depicts the percentiles of one distribution relative to the other:

$$p_2 = F\left(G^{-1}(p_1)\right),\ 0 \leq p_1 \leq 1, \tag{1}$$

whereby $G^{-1}(y) = \inf \{x : G(x) \geq y\}$. An example of such a PP-plot is given in Panel a of Figure 1. Clearly PP-plots are mappings from $[0, 1]$ onto $[0, 1]$ and depict the correspondence between the two underlying distributions in probability space.

4

2.1 HM index: definition

As the PP-plot coincides with the diagonal if, and only if, the two underlying distributions F and G are identical, an obvious measure to characterize PP-plots is the absolute value of the total deviation between the diagonal and the PP-plot. This notion is much in the same spirit as the Gini coefficient for Lorenz curves. To normalize this measure to an index value between 0 and 1, the particular surface is multiplied by 2, that is:

Definition 1

$$HM(F,G) \equiv 2 \int_0^1 \left| p - F\left(G^{-1}(p)\right) \right| dp. \tag{2}$$

Indeed, the HM index corresponds to the shaded area multiplied by 2 in Panel b of Figure 1.

2.2 HM index: properties

Observe a number of properties of the HM index.

Property P1 (equality): $HM(F,G) = 0 \iff \forall\ q \in [a,b] : F(q) = G(q)$.

As $HM(F,G) = 0$ implies that $p = F\left(G^{-1}(p)\right)\ \forall\ p \in [0,1]$, it follows that $F = G$; likewise, $F(q) = G(q)$ implies that $F^{-1}(G(q)) = q$, from which an HM index of 0 follows.

Property P2 (symmetry): $HM(F,G) = HM(G,F)$.

As the PP-plot of F taking G as the basis is the reflection of the PP-plot around the diagonal of G taking F as the basis, either yields the same value for the HM index.

Property P3 (range): $HM(F,G) \in [0,1]$.

Since a PP-plot is an increasing function on its domain, the maximum deviation with the diagonal is obtained if it never crosses in the interior and approaches either of the coordinates (0,1) or (1,0), in which case the absolute surface area between the diagonal and the PP-plot is $\frac{1}{2}$ (this maximum value can also be obtained if the PP-plot crosses the diagonal only once).

3   The sample Harmonic Mass index

For using HM index (2) in practise its sample counterpart needs to be considered. Let $X_n = \{x_1, ..., x_n\}$ be n realizations of finitely discrete random

Month	1990	1993
January	4070	4114
February	4204	3944
March	3885	3814
April	3866	3813
May	3808	3836
June	3854	3757
July	3762	3824
August	3786	3751
September	3764	3809
October	3872	3896
November	3926	3818
December	4292	4406

Table 1: Maximum water levels of the river Meuse at Borgharen Dorp, in millimeters, measured on every last day of every month. Source: Koninklijk Nederlands Meteorologisch Instituut, personal correspondence.

variable X, yielding as empirical CDF F_n, and let $Y_m = \{y_1, ..., y_m\}$ be m realizations of finitely discrete random variable Y, with empirical CDF G_m. The ordered set of values which either X or Y assumes is denoted by $c_1, ...c_k$, whereby $k \leq n + m$. In addition let c_0 and c_{k+1} denote $-\infty$ and ∞ respectively and define $C \equiv \{c_0, ..., c_{k+1}\}$. The horizontal coordinates of the discrete sample PP-plot are then defined by $P[X \leq c_i] \; \forall \; c_i \in C$ while the vertical coordinates are given by $P[Y \leq c_i] \; \forall \; c_i \in C$ (Bamber, 1975).

Consider for example the water level of the river Meuse as it enters The Netherlands at Borgharen Dorp in 1990 and 1993. In 1993 the Southern part of Holland was plagued by severe floods and it is of interest to know whether the entire year 1993 was exceptional. Table 1 contains the maximum water levels in millimeters for both 1990 and 1993 recorded on each last day of every month. The resulting discrete sample PP-plot is depicted in Panel a of Figure 2.

To compare 1990 with 1993 the Mww U statistic could be computed:

$$Mww(X_n, Y_m) \equiv \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} \delta_{ij}, \qquad (3)$$

where

$$\delta_{ij} = \begin{cases} 0, & X_i < Y_j, \\ \frac{1}{2}, & X_i = Y_j, \quad i = 1, ..., n, \; j = 1, ..., m. \\ 1, & X_i > Y_j, \end{cases}$$
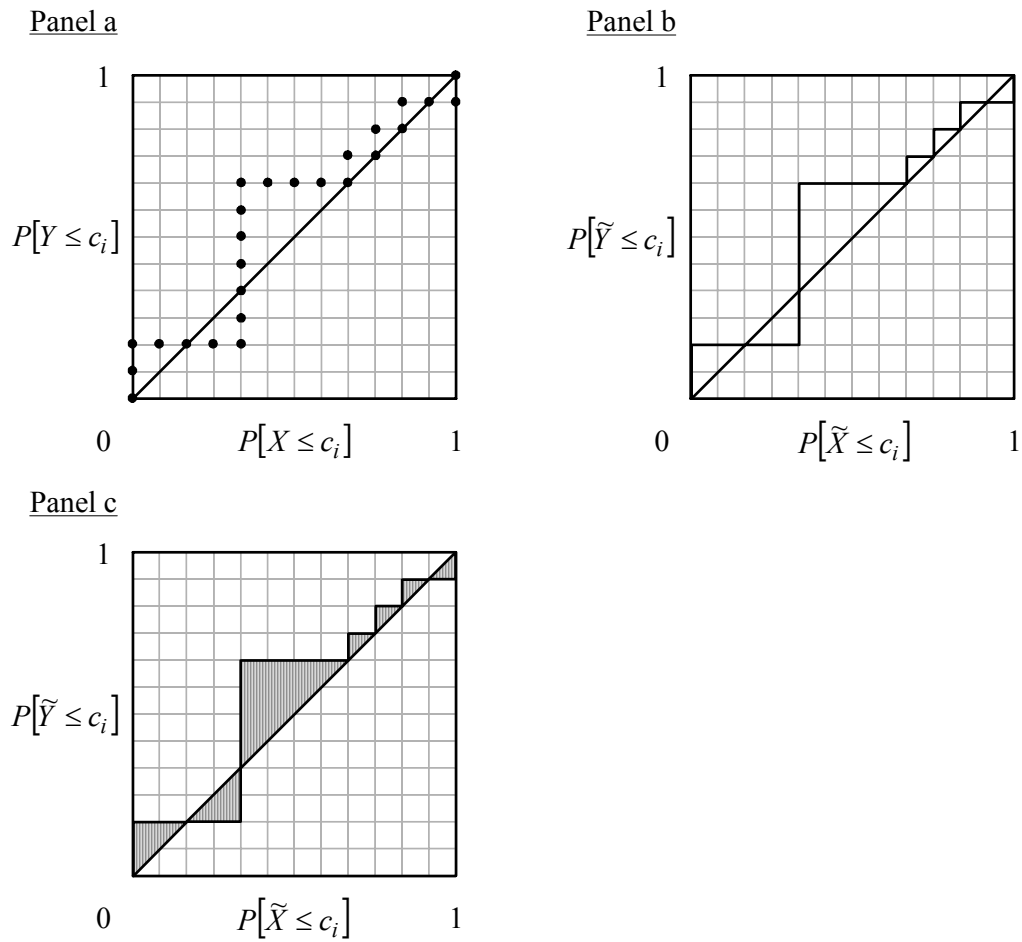
Figure 2: Discrete sample PP-plot for entering heights of the river Meuse into The Netherlands at Borgharen Dorp, 1990 versus 1993 (Panel a), the corresponding sample PP-plot (Panel b), and the concomitant sample HM index (Panel c).

In case X refers to the 1990 water levels, (3) equals $\frac{7}{16}$ and the hypothesis that $P[X \leq c_i] > P[Y \leq c_i]$ is rejected. The reverse hypothesis is rejected as well considering the value of $\frac{9}{16}$ for the Mww U statistic. Yet, this leaves undecided whether sample years 1990 and 1993 are different in a probabilistic sense regarding the entering heights of the river Meuse into The Netherlands.

3.1   Sample HM index: definition

To establish if $P[X \leq c_i] \neq P[Y \leq c_i]$, the sample HM index can be considered. It is derived from the continuous analogue of the discrete sample PP-plot – referred to as the sample PP-plot – which is obtained by uniformly distributing the probability that is concentrated in any point c_i over the interval from c_{i-1} to c_i (Girling, 2000). If the function α is defined as $\alpha(c_k) = c_{k-1}$, and if v and ω are independent random variables that are uniformly distributed over the unit interval, the continuous analogues of X and Y follow from:

$$
\begin{aligned}
\widetilde{X} &= vX + (1-v)\alpha(X), \\
\widetilde{Y} &= \omega Y + (1-\omega)\alpha(Y).
\end{aligned}
$$

The sample PP-plot is thus obtained by connecting the points of the discrete sample PP-plot through straight lines as in Panel b of Figure 2 (Bamber, 1975). The empirical CDFs of $\widetilde{X}$ and $\widetilde{Y}$ are denoted by $\widetilde{F}_n$ and $\widetilde{G}_m$ respectively. These are continuous functions with a finite number of steps.

   The sample HM index is then defined as:

Definition 2

$$
HM^S\left(\widetilde{F}_n, \widetilde{G}_m\right) \equiv 2 \int_0^1 \left| p - \widetilde{F}_n\left(\widetilde{G}_m^{-1}(p)\right) \right| dp. \tag{4}
$$

In Panel c of Figure 2 the sample HM index corresponds to the shaded area and equals $\frac{7}{36}$. In this case the value of the sample HM index can be obtained by a straightforward geometric argument. In Section 3.2 below a simple rule is specified that yields the exact value of the sample HM index when both samples are of equal size and absent within-sample and between-samples ties (as is the case here, see Table 1). For this special case also the exact distribution of the HM index is derived under the hypothesis that both samples are drawn from the same distribution (Section 4). Considering this distribution indicates that this hypothesis cannot be rejected; statistically
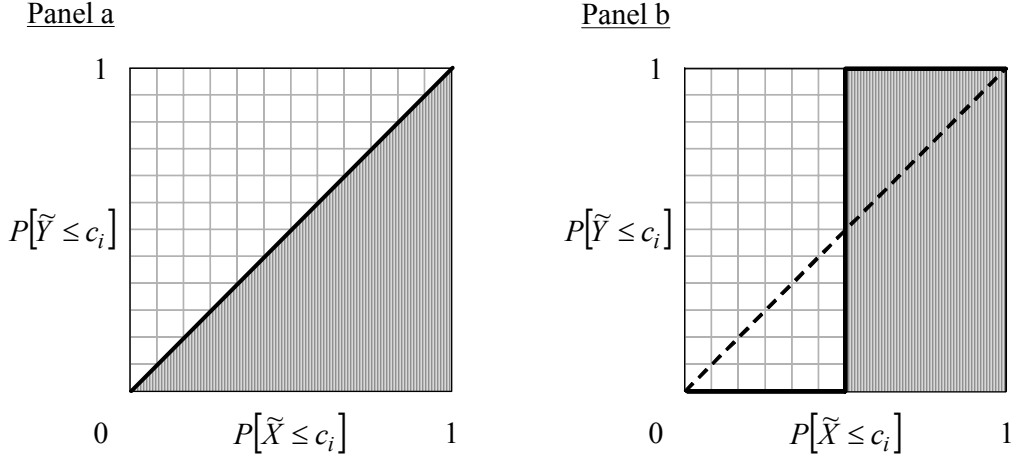
Figure 3: The relation between the Mww U statistic and sample PP-plots; a value of $\frac{1}{2}$ is obtained for two identical samples (Panel a), and in case all entries between the two samples differ due to, for instance, a mean-preserving increase in spread of one sample (Panel b).

1990 and 1993 were not different regarding the entry height of the river Meuse into The Netherlands.

The notion that the Mww U statistic should not be used to assess if two samples are different in a probabilistic sense also follows from its relation to the sample PP-plot. Bamber (1975) shows that the value of the Mww U statistic (3) corresponds to the area below the sample PP-plot (other versions of the Mww U statistic can be represented in a similar way, see Girling, 2000). Obviously, in case $X_n = Y_m$ the Mww U statistic equals $\frac{1}{2}$ (Panel a of Figure 3), a value that is also obtained for the sample PP-plot depicted in Panel b of Figure 3. This latter plot depicts what could happen due to a mean-preserving increase in spread of one of the underlying samples.

Finally, note that the sample PP-plot converges point-wise to its theoretical counterpart with probability one when $n, m \to \infty$ (Mushkudiani, 2000). The following results are then immediate:

Proposition 1 $lim_{n,m\to\infty} HM^S\left(\widetilde{F}_n, \widetilde{G}_m\right) = HM(F,G)$.

This consistency of the sample HM index implies in particular:

Corollary 1 $lim_{n,m\to\infty} HM^S\left(\widetilde{F}_n, \widetilde{G}_m\right) = 0 \iff \forall\, q \in [a,b] : F(q) = G(q)$.

Accordingly, the sample HM index converges to zero if, and only if, the two samples are drawn from the same distribution. This makes it an appropriate statistic for testing just that. Indeed, in Section 4 below, the distribution of the sample HM index is derived under the assumption that both samples are drawn from the same distribution for a set of CDFs that is a subset of Ξ_1.

3.2 Sample HM index: computation

To approximate sample HM index (4) numerous computing packages are available.[3] In special cases the sample HM index can be determined without approximation error.

First, an exact identification of the discrete sample PP-plot is possible when the following two assumptions hold:

Assumption A1: $rank(X_n) = n$, $rank(Y_m) = m$.

Assumption A2: $rank(X_n, Y_m) = n + m$.

Assumption *A1* implies that no ties are present within either sample; assumption *A2* precludes ties between samples.

Let $R(\cdot)$ denote the rank of an element from the joint, ordered sample $\{X_n, Y_m\}$ whereby the largest sample value receives rank 1. The coordinates of the discrete sample PP-plot then follow (see Appendix A for the proof):

Lemma 1 A1-A2 *hold* $\implies$

$$\left(t^X_{1,i}, t^X_{2,i}\right) = \left(\frac{i}{n}, \frac{n+m+1-i-R(x_i)}{m}\right), \quad i = 1, ..., n,$$
$$\left(t^Y_{1,j}, t^Y_{2,j}\right) = \left(\frac{n+m+1-j-R(y_j)}{n}, \frac{j}{m}\right), \quad j = 1, ..., m.$$

Together with the origin (0,0) the coordinates in Lemma 1 provide the exact discrete sample PP-plot (as in Figure 4 for example). Obviously, reversal of X_n and Y_m results in the reflection of the discrete sample along the diagonal.

Second, from this sample PP-plot the exact concomitant sample HM index can be computed, using only the "X-coordinates" (the proof is in Appendix A):

Lemma 2 A1-A2 *hold* $\implies$

$$HM^S\left(\widetilde{F}_n, \widetilde{G}_n\right) = \sum_{i=1}^{n} \left\{ \begin{array}{c} I_i\left[\left(t^X_{1,i-1} - t^X_{2,i}\right)^2 + \left(t^X_{1,i} - t^X_{2,i}\right)^2\right] \\ + (1 - I_i)\left(t^X_{1,i} - t^X_{1,i-1}\right)\left|2t^X_{2,i} - t^X_{1,i} - t^X_{1,i-1}\right| \end{array} \right\},$$

___

[3] A GAUSS routine that computes sample HM index (3) using an algorithm based on Chotikapanich and Griffiths (2001) is available upon request.
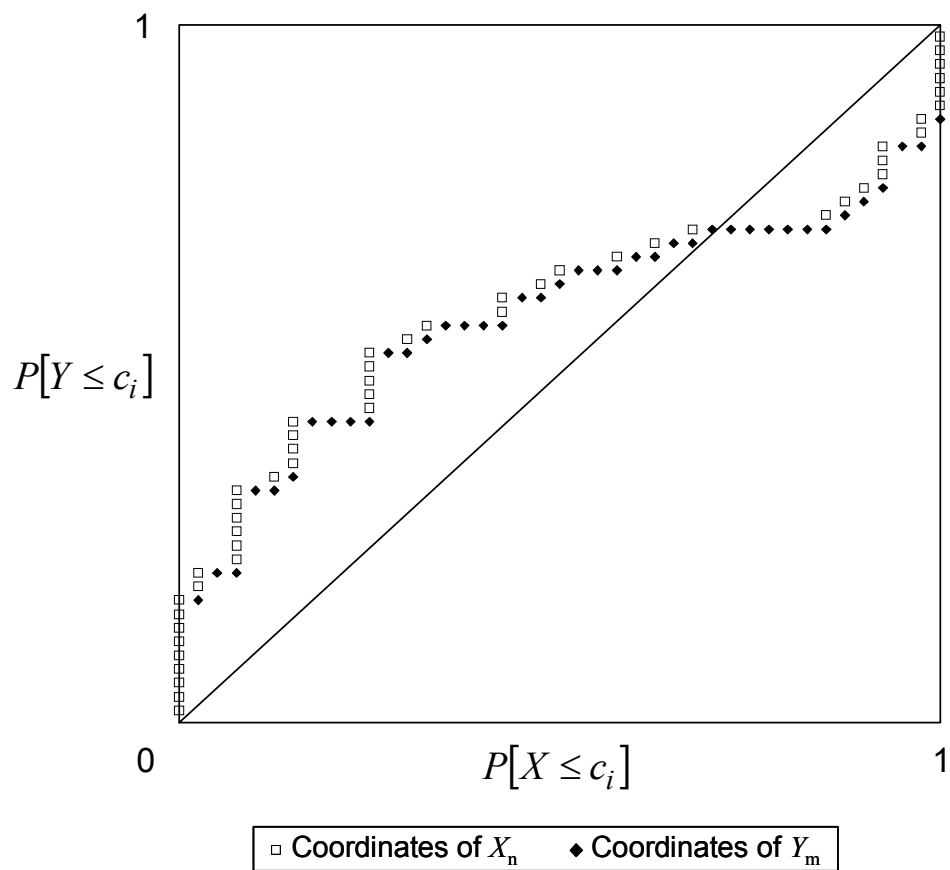
Figure 4: Discrete sample PP-plot based on Lemma 1 for $X \sim N(0, 1)$, and $Y \sim N(0, 4)$, with $n = 50$, and $m = 40$.

11

whereby

$$I_i = \begin{cases} 1 & if\ t^X_{1,i-1} < t^X_{2,i} < t^X_{1,i} \\ 0 & otherwise \end{cases} .$$

A simpler form of the exact sample HM index is available if in addition the following assumption applies:

Assumption A3: $n = m$.

Section 4.2 below contains two suggestions for adjusting samples to be of equal size. For computing the sample HM index this is useful as (Appendix A contains the proof):

Lemma 3 A1-A3 *hold* $\Longrightarrow$

$$HM^S\left(\widetilde{F}_n, \widetilde{G}_n\right) \equiv \frac{1}{n}\sum_{i=1}^{n}\left\{\left|t^X_{1,i} - t^X_{2,i}\right| + \left|t^Y_{1,i} - t^Y_{2,i}\right|\right\}.$$

Sample HM index (4) is thus readily computed using Lemmata 1 and 2 provided that *A1-A2* hold. If in addition *A3* holds, the computationally less burdensome formulation in Lemma 3 can be used.

4 Hypotheses testing

To test statistically whether two samples are drawn from the same distribution, the probability density function (pdf) of the sample HM index is to be retrieved under this hypothesis. Because the number of possible sample HM index values for any two samples X_n and Y_m quickly becomes countably infinite with increasing sample sizes, a general characterization of this pdf cannot be provided. For a subset of Ξ_1 this is possible however.

In particular, let $\Xi_2 = \{F|\,\forall x,\, h \in \mathbb{R}: \lim_{x \to -\infty}F(x) = 0,\, \lim_{x \to \infty}F(x) = 1,\, \lim_{h \longrightarrow 0}F(x + h) = F(x),$ and $a < b \Longrightarrow F(a) < F(b),$ for $F(a), F(b) \in (0,1)\}$. Note that $\Xi_2 \subset \Xi_1$, that functions belonging to Ξ_2 are continuous and strictly increasing CDFs on their support, and that mass points are absent. Hence, the probability that assumptions *A1 – A2* are met for samples drawn from some $F \in \Xi_2$ is one.

Given any two samples X_n and Y_n, respectively drawn from (unknown) distributions $F, G \in \Xi_2$, the hypothesis of interest is:

H_0: $F = G$.

4.1 Symmetric samples

For deriving the pdf of $HM^S\left(\widetilde{F}_n, \widetilde{G}_n\right)$ under H_0 for any two samples of size $n \in \mathbb{N}$, the following notation needs to be introduced:

- $\Theta(n) \equiv 1 + n(n-1)/2$, the number of possible distinct values of the sample HM index;

- $HM_n^S \equiv \left(HM_n^S(j)\right)$, the vector containing all possible distinct sample HM index values, whereby $HM_n^S(j) \equiv 1 - 2j/n^2$, $j = 0, ..., n(n-1)/2$;

- $\Omega(n) \equiv 2 \prod_{i=0}^{n-2} \frac{[2(n-i)-1]}{(n-i)}$, half the number of possible distinct discrete sample PP-plots, whereby $\Omega(1) \equiv 1$;

- $B(n) \equiv (\Delta_n | 2\Upsilon_n)$, whereby Δ_n is the upper triangular unit matrix of size n and Υ_n the unit vector of size n;

- $A(n+1) \equiv M(n)B(n)$, whereby $M(n) \equiv 1$ and $M(n+1) = (m_{jk}(n + 1))$ with

$$m_{jk}(n+1) = \begin{cases} a_{(j-k+1)k}(n+1), & j = k, ..., \Theta(n) + k, \ k = 1, ..., n+1, \\ 0 & otherwise. \end{cases}$$

In Appendix A the following proposition is proved.

Proposition 2 *Let $X \sim F \in \Xi_2$, and $Y \sim G \in \Xi_2$. Under H_0, given any two samples X_n and Y_n with, respectively, concomitant empirical CDFs $\widetilde{F}_n$ and $\widetilde{G}_n$, for any symmetric sample size $n \in \mathbb{N}$ the following holds:*

$$h\left(HM_n^S(i)\right) = P\left[HM^S\left(\widetilde{F}_n, \widetilde{G}_n\right) = HM_n^S(i)\right] = d_i(n),$$

whereby

$$D(n) = (d_i(n)); \ d_i(n) = \frac{M(n)\Upsilon_n}{\Omega(n)}, \ i = 1, ..., \Theta(n).$$

The pdf $h(x)$ of the sample HM index is given in Table 2 for $n = 1, ..., 6$. The number of possible distinct outcomes of the HM index increases quadratically with n and is, for example, equal to 191 for $n = 20$, and 4,951 for $n = 100$. The number of possible distinct discrete sample PP-plots increases much more rapidly and equals 68,923,264,410 for $n = 20$, and approximately equals 4.52743×10^{58} for $n = 100$. Also, and in line with corollary 1, the probability mass moves towards lower values of the sample HM index with increasing sample sizes, as illustrated in Figure 5.

13

$HM_n^S(i)$	$h\left(HM_n^S(i)\right)$	$HM_n^S(i)$	$h\left(HM_n^S(i)\right)$
$n=1$		$n=5$ (continued)	
1	1	15/25	7/126
$n=2$		13/25	13/126
1	1/3	11/25	20/126
2/4	2/3	9/25	24/126
$n=3$		7/25	32/126
1	1/10	5/25	16/126
7/9	1/10	$n=6$	
5/9	4/10	1	1/462
3/9	4/10	34/36	1/462
$n=4$		32/36	2/462
1	1/35	30/36	3/462
14/16	1/35	28/36	5/462
12/16	2/35	26/36	9/462
10/16	5/35	24/36	11/462
8/16	6/35	22/36	15/462
6/16	12/35	20/36	22/462
4/16	8/35	18/36	34/462
$n=5$		16/36	37/462
1	1/126	14/36	58/462
23/25	1/126	12/36	72/462
21/25	2/126	10/36	80/462
19/25	3/126	8/36	80/462
17/25	7/126	6/36	32/462

Table 2: Pdf of the sample HM index under H0 for samples of equal size, for n = 1,...,6.
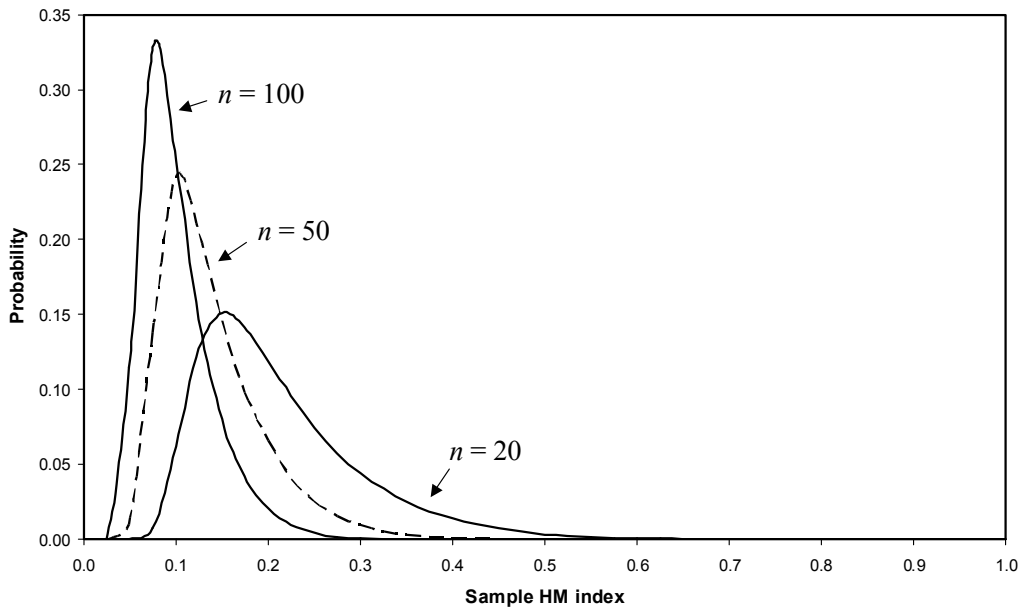
Figure 5: Pdf of the sample HM index for different symmetric sample sizes.

Appendix B lists the first four moments of $h(x)$, four critical percentiles, and the variance of the distribution for $n = 3, ..., 200$ and $n = 250, 300, 350$.[4] This allows for obvious hypothesis testing for symmetric samples of modest size. For larger samples a rather accurate approximation of the various critical percentiles is available. Table 3 contains these approximations which are ruled by the linear regression $\ln(cp) = \beta_0 + \beta_1 \ln(n)$, where cp refers to the critical percentile, $cp = \{0.90, 0.95, 0.975, 0.99\}$.[5] These approximations are accurate indeed as for $n \geq 100$ the absolute difference between the actual and predicted critical percentile is less than 10^{-3} (considering the goodness-of-fit of the various regressions this should not come as a surprise).[6] For instance, for $n = 1000$ the 99th percentile of the sample HM index CDF approximately equals 0.0673.

---

[4]The current state of our technology – Intel Pentium 4 processor with a CPU of 3.06 GHz and 1 GB RAM – does not allow for an assessment of larger sample sizes.

[5]All regressions in Table 3 are based on the entries in Appendix B for sample sizes $n \geq 100$.

[6]The estimated slopes are remarkably similar for the various critical percentiles; a pooled regression yields as slope estimate -0.5002, and intercepts of, respectively, 0.3472, 0.5003, 0.6236, 0.7550. It does not, however, improve the prediction performance of the regression.

15

Critical percentile: $\ln(cp) = \beta_0 + \beta_1 \ln(n)$				
Percentile	90	95	97.5	99
β_0	0.3503	0.5018	0.6230	0.7508
	(0.0008)	(0.0006)	(0.0005)	(0.0007)
β_1	-0.5008	-0.5005	-0.5001	-0.4994
	(0.0002)	(0.0001)	(0.0001)	(0.0001)
R^2	0.9999	0.9999	0.9999	0.9999

Table 3: Rule-of-thumb for determining various critical percentiles of the CDF of the sample HM index; heteroskedasticity-consistent standard errors in parentheses (White, 1980).

4.2 Asymmetric samples

In practise different samples easily have different sizes, which still allows for computation of sample HM index (4) though (see also footnote 3). To use it for hypotheses testing requires either an approximation of its distribution, or that the problem is converted towards the symmetric sample case.

4.2.1 Monte Carlo

The distribution of the sample HM index under H_0 can always be obtained through simulation. Table 4 contains simulated critical percentiles of the distribution of sample HM index (4) for various asymmetric sample pairs. Each simulated distribution is based on 100,000 runs and the underlying samples are drawn from a standard normal distribution. Note that the critical values of the two symmetric sample pairs ($n = 50$ and $n = 100$) are close approximations of their true values (see Appendix B).

4.2.2 Symmetric approximation

Alternatively an educated guess is formulated regarding the relevant critical percentiles. One such guess would be the critical percentiles of the symmetric distribution with $\overline{n} = (n + m)/2$ observations (appropriately adjusted in case $\overline{n}$ would not be an integer). Considering the entries in Table 4 and in Appendix B suggests that this procedure yields accurate critical percentile values in case the smallest sample is not 'too small' (say, larger than 30), and the relative difference in sample size is not 'too large' (say, less than 10%). For instance, for $n = 90$ and $m = 100$ the critical values are in Table 4, which compare rather accurately with the theoretical values for symmetric sample size $\overline{n} = 95$ in Appendix B.

	$n_1 = 50$				$n_1 = 100$			
n_2	z_{90}	z_{95}	$z_{97.5}$	z_{99}	z_{90}	z_{95}	$z_{97.5}$	z_{99}
10	0.3480	0.4040	0.4520	0.5120	0.3327	0.3864	0.4353	0.4916
20	0.2651	0.3091	0.3480	0.3954	0.2443	0.2837	0.3203	0.3663
30	0.2320	0.2693	0.3047	0.3447	0.2087	0.2424	0.2738	0.3113
40	0.2127	0.2483	0.2806	0.3187	0.1871	0.2180	0.2458	0.2806
50	0.2000	0.2336	0.2648	0.2992	0.1731	0.2021	0.2281	0.2604
60	0.1919	0.2236	0.2526	0.2868	0.1628	0.1900	0.2145	0.2467
70	0.1856	0.2161	0.2450	0.2784	0.1559	0.1816	0.2058	0.2331
80	0.1808	0.2096	0.2374	0.2700	0.1501	0.1752	0.1975	0.2254
90	0.1759	0.2055	0.2328	0.2641	0.1456	0.1692	0.1913	0.2188
100	0.1731	0.2015	0.2284	0.2605	0.1418	0.1648	0.1868	0.2122

Table 4: Simulated critical percentiles z of the sample HM index CDF for various asymmetric sample sizes.

4.2.3 (Random) selection

Because simulation might be considered to be too tedious[7] and the symmetric approximation too much off the mark, alternative solutions to the violation of $A3$ equalize the number of observations in both samples. For instance, the number of observations in the larger sample can be reduced to match that of the smaller sample. Even though this procedure discards valuable information and it is not immediate which observations are to be removed, especially for large sample sizes this can be an effective solution. To tackle the problem of observation selection various subsamples from the larger sample could be considered to qualify the robustness of the computed sample HM index. Alternatively an additional indicator is available (industry or country code, sex, age, etc.) along which an intersection of samples can be composed.

Comparing the symmetric sample size pairs in Table 4 with the asymmetric pairs suggests the order of the Type-I error of this procedure. Although for individual cases the effect on the sample HM index of removing observations from the larger sample is ambiguous, on average it has the effect that the computed HM index is too large. Hence, H_0 is rejected too often and a sufficient condition emerges; if H_0 is not rejected for samples that are symmetric by construction, it most likely would not be rejected for the original, asymmetric setting.

---

[7]Yet, the GAUSS routine that simulates the distribution of the sample HM index for sample pairs of any size is available upon request.

	Percentage of probability mass			
Pollution percentage	90%	95%	97.5%	99%
1%	0.9018	0.9506	0.9755	0.9902
2%	0.9001	0.9497	0.9752	0.9903
3%	0.8972	0.9491	0.9742	0.9894
4%	0.8964	0.9481	0.9735	0.9895
5%	0.8921	0.9474	0.9738	0.9897
10%	0.8685	0.9357	0.9692	0.9879
15%	0.8214	0.9157	0.9603	0.9848
20%	0.7457	0.8790	0.9446	0.9798
25%	0.6378	0.8212	0.9171	0.9706
50%	0.0586	0.2053	0.4343	0.7215

Table 5: Fraction of HM index values corresponding to the respective percentiles of the HM index CDF for varying degrees of data contamination.

4.2.4 Interpolation

Alternatively the number of observations in the smaller sample is increased to match the size of the larger sample. Various methods are available for series fitting to sample data (see, e.g. Cabral and Mata, 2003). These fitted series then allow for an alignment of the two samples sizes.

5 Robustness

For assessing the robustness of the HM test two situations are considered: (i) samples drawn from the same distribution whereby a varying proportion of the data is corrupted, and (ii) samples drawn from different distributions absent outliers.

5.1 Type I errors

Table 5 contains the fractions of the HM index values being less than or equal to the true HM index value corresponding to the respective percentiles. These entries are based on 100,000 runs with sample size $n = 100$. Both X_n and Y_n are drawn from the standard normal distribution with, however, a varying percentage of the observations in Y_n being drawn from $N(0, 100)$. This is a typical procedure to create outliers (see e.g. Wagenvoort and Waldman, 2002). Note that for fractions of data corruption beyond 50 percent the identification of the true distribution of Y becomes blurred.

90%			
	$\ln(X) \sim N(0,1)$	$X \sim UNIF(0,3)$	$X \sim EXP(1)$
$\ln(Y) \sim N(0,3)$	0.9007	0.1610	0.1086
$Y \sim UNIF(0,1)$	0.1610	0.9005	0.0009
$Y \sim EXP(1)$	0.1086	0.0009	0.9009
95%			
	$\ln(X) \sim N(0,1)$	$X \sim UNIF(0,3)$	$X \sim EXP(1)$
$\ln(Y) \sim N(0,1)$	0.9502	0.2868	0.1772
$Y \sim UNIF(0,3)$	0.2868	0.9510	0.0023
$Y \sim EXP(1)$	0.1772	0.0023	0.9506
97.5%			
	$\ln(X) \sim N(0,1)$	$X \sim UNIF(0,3)$	$X \sim EXP(1)$
$\ln(Y) \sim N(0,1)$	0.9755	0.4220	0.2580
$Y \sim UNIF(0,3)$	0.4220	0.9751	0.0056
$Y \sim EXP(1)$	0.2580	0.0056	0.9753
99%			
	$\ln(X) \sim N(0,1)$	$X \sim UNIF(0,3)$	$X \sim EXP(1)$
$\ln(Y) \sim N(0,1)$	0.9903	0.5886	0.3759
$Y \sim UNIF(0,3)$	0.5886	0.9903	0.0139
$Y \sim EXP(1)$	0.3759	0.0139	0.9899

Table 6: Fraction of HM index values corresponding to the respective percentiles of the HM index CDF for samples drawn from distributions.

On average, data contamination increases the value of the sample HM index. For fractions of data contamination up to five percent the HM test remains quite reliable; approximation errors of the critical percentiles are well within a one percentage point. On the other hand, using the HM test with the lower critical percentiles is especially prone to Type-I errors if more than 10 percent of the data are corrupted.

5.2 Type II errors

Alternatively the two underlying distributions differ by construction. As the number of comparisons is unlimited, all that can be provided is an illustration of Type-II errors for a particular set of distributions. Our choice of distributions is ruled by their commonality. The distribution-specific parameters are chosen such that the distributions mimic each other.

In Table 6 again the true HM index value for the various critical percentiles is compared with that of 100,000 runs for each of the distribution

pairs (with, again, $n = 100$). In all cases, the fraction of computed HM index values that is below the HM index value corresponding to a specific percentile of the HM index CDF, is much lower than the true percentage. Although these simulations do not provide an exhaustive assessment of the power of the HM test, they do suggest that the test is capable of discerning different, commonly used distributions.

6 An application

Using the sample HM index is of particular use for tracing out the development of distributions over time because it avoids clustering observations into a discrete number of cells, as is necessary for instance when applying traditional Markov analyses. Consider for example the distribution over all 4-digit sectors of the UK regarding their value of the Balassa index – which measures the comparative advantage of a sector – from 1970 through 1997. The Balassa index is defined as a sectors' national export share as a fraction of the sectors' export share in world exports. It is thus a positive, real number whereby index values above one refer to a "revealed comparative advantage" (see Balassa, 1965).[8] On average, throughout the sample period the UK experiences a comparative advantage for 36.7% of all its sectors where a maximum index value of 10.39 is recorded in 1974 for "Spirits; liqueurs, and other spirituous beverages (excluding wine, cider, Perry mead, ale, stout and porter)".

As comparative advantage tends to be sticky, a five year comparison lag is used to assess the development of the distribution of Balassa indices over time. Figure 6 contains the sample PP-plots of the Balassa index distribution for the UK in 1978 versus 1983 (Panel a), and 1988 versus 1993 (panel b) including the concomitant values of the sample HM index. The Balassa indices are calculated using the Feenstra (2000) data. The sample HM index values for all year-by-year comparisons are in Table 7 together with the appropriate critical percentiles at the 5 percent significance level, denoted by $z_{95}\left(HM^S\right)$. The latter are obtained using the rule-of-thumb specified in Table 3 and can thus differ per comparison as the underlying samples vary in size. The year-by-year comparisons are restricted to the subset of sectors for which a strictly positive Balassa index is recorded in both years; symmetry in

---

[8]Hillman (1980) identifies a sufficient condition for the index value to measure comparative advantage proper in that an increase in a sectors' exports yields an increase in the related Balassa index. The analysis here is restricted to observations meeting this Hillman condition, which boils down to dismissing some 0.01% of all observations that together represent 0.5% of the value of total trade.

Sample years	HM^S	$z_{95}\left(HM^S\right)$	MI
1970, 1975	0.0644	0.0823	0.4219
1971, 1976	0.0681	0.0823	0.4125
1972, 1977	0.0668	0.0823	0.4219
1973, 1978	0.0328	0.0826	0.5315
1974, 1979	0.1147	0.0825	0.5282
1975, 1980	0.1145	0.0848	0.4769
1976, 1981	0.0955	0.0845	0.5033
1977, 1982	0.1273	0.0844	0.5180
1978, 1983	0.1364	0.0839	0.4156
1979, 1984	0.1241	0.0838	0.4141
1980, 1985	0.0979	0.0838	0.4433
1981, 1986	0.0682	0.0840	0.4497
1982, 1987	0.0314	0.0837	0.4389
1983, 1988	0.0229	0.0839	0.4448
1984, 1989	0.0301	0.0840	0.4562
1985, 1990	0.0225	0.0845	0.4572
1986, 1991	0.0312	0.0844	0.4785
1987, 1992	0.0259	0.0841	0.4542
1988, 1993	0.0286	0.0843	0.5001
1989, 1994	0.0281	0.0844	0.3967
1990, 1995	0.0365	0.0844	0.4425
1991, 1996	0.0303	0.0841	0.3889
1992, 1997	0.0263	0.0843	0.4152

Table 7: Year-by-year comparisons of the distribution of the Balassa index in the UK, using the sample HM index (second column) and a mobility index based on Markov matrices (fourth column).
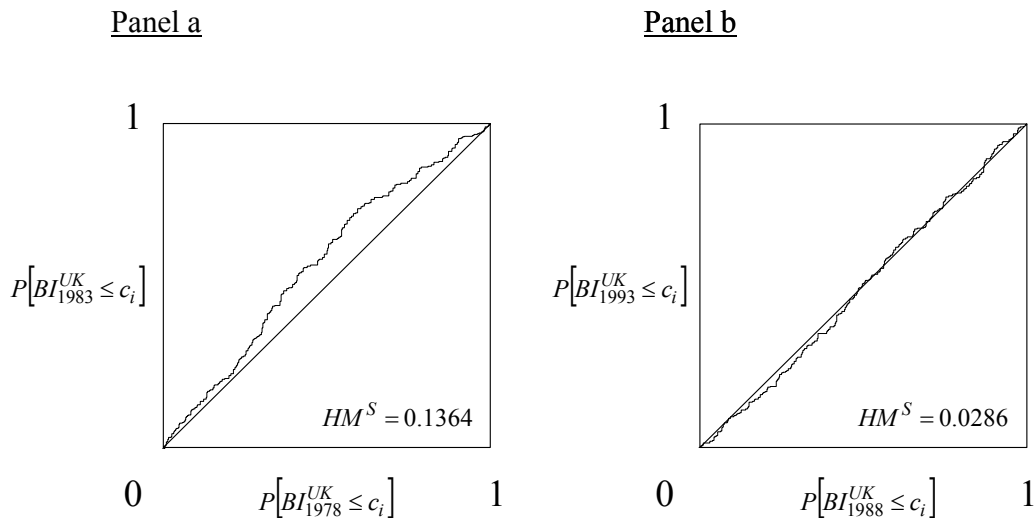
Figure 6: Sample PP-plots based on the Balassa index distribution of some 400 4-digit sectors in the UK for the years 1978 versus 1983 (Panel a), and 1988 versus 1993 (Panel b).

sample size is thus obtained using the selection method discussed in Section 4.2.

For the UK the years 1974 – 1980 stand out in that the distribution of the Balassa index differs significantly from the same distribution five years later. An in-depth analysis of the UK economy goes beyond the scope of this paper but do observe that the identified period coincides with the first and second oil crisis. Together with the particularities of the labour unions in the UK at the time would account for the identified pattern (indeed, the same procedure applied to other European countries does not identify the same years as particularly volatile regarding revealed comparative advantage).

Alternatively the year-to-year dynamics of the process are captured by Markov matrices $\Lambda_{i,j}$, and the information contained in each of these transition matrices is converted into a single number using an appropriate norm.[9] The fourth column of Table 7 contains the so obtained mobility indices MI, whereby the respective Markov matrices are based on quintiles and the mobility indices defined as (Geweke *et al.*, 1986): $MI = (n - tr(\Lambda_{i,j}))/(n-1)$.[10]

---

[9]This alternative is discussed as it is typically used in the literature for describing the dynamics of empirical distributions. However, Markov matrices capture the volatility of the rank of observations while the HM index as applied here captures the volatility of the shape of the various sample distributions.

[10]There is an element of discreteness here in the sense that the number of cells in the

Considering then these mobility indices does not identify the period 1974 – 1980 as being special. Some of the years in this time span could be labelled particularly volatile if that coincides with $MI \geq \frac{1}{2}$, but that would then also apply to 1988. However, and relatedly, no statistical test is available for comparing mobility indices.[11] Indeed, qualitatively the traditional Markov approach does not identify proper the years that are especially volatile regarding the change in specialization patterns of the UK economy; quantitatively no test is available to assess the statistical significance of the related mobility indices.

7   Conclusions

In this paper the graphical information of PP-plots is translated into a single index number. As PP-plots characterize the extent to which the two underlying probability masses are in harmony, the index is labelled the Harmonic Mass index. The sample version of the HM index, which is based on sample PP-plots, is shown to be consistent. For samples of equal size and absent within-sample and between-sample ties, elegant rules are derived for computing the sample HM index exactly.

For a wide class of CDFs the exact distribution of the sample HM index is derived, under the hypothesis that the two underlying samples are drawn from the same distribution. Accordingly, a new, nonparametric, distribution-free statistic is developed for testing exactly that.

Because of the burdensome computations that are involved to determine critical percentiles of the statistics' CDF, these are provided for samples that include up to 350 entries. The current state of available techniques does not allow for an assessment of larger sample sizes. However, an accurate rule-of-thumb is presented for determining these critical values for any, symmetric sample size. Also, as for some applications the sample distributions that are to be compared are of unequal size, four methods are discussed that translate any asymmetric setting towards the symmetric case.

Simulations indicate that the HM test is relatively unaffected by outliers if these corrupt up to 10 percent of the data. For assessing the type-II error of the HM test three commonly observed distributions are considered:

Markov matrices are chosen *ad hoc*. Varying this number yields different values for MI, but qualitatively leaves the mobility indices unaltered.

[11] A test for comparing Markov matrices does exist (Anderson and Goodman (1957); for an application see Proudman and Redding, 2000). However, this particular test does not specify what to do in case matrix cells are empty, a situation that is quite often encountered in practise.

log-normal, uniform, and exponential, whereby the distributions are parameterized such that they mimic each other. The simulations suggest that the HM test is indeed capable of discerning these commonly used distributions.

Finally, an application regarding the dynamic development of revealed comparative advantage not only stresses the usefulness of the HM index, it also illustrates the use of one of the methods to compare samples of different size.

References

[1] Anderson, T. W. and L. A. Goodman (1957), "Statistical inferences about Markov chains", *The Annals of Mathematical Statistics*, **28**(1): 89 – 110.

[2] Balassa, B. (1965), "Trade liberalization and 'revealed' comparative advantage", *The Manchester School of Economic and Social Studies*, **33**: 92 – 123.

[3] Bamber, D. (1975), "The area above the ordinal dominance graph and the area below the receiver operating characteristic graph", *Journal of Mathematical Psychology*, **12**: 387 – 415.

[4] Cabral, L. M. B and J. Mata (2003), "On the evolution of the firm size distribution: facts and theory", *American Economic Review*, **93**(4): 1075 – 1090.

[5] Chotikapanich, D. and W. Griffiths (2001), "On calculation of the extended Gini coefficient", *Review of Income and Wealth*, **47**(4): 541 – 547.

[6] Feenstra, R. C. (2000), *World trade flows; 1980 – 1997*, Institute of Governmental Affairs, University of California, Davis.

[7] Geweke, J., R.C. Marshall and G.A. Zarkin (1986), "Mobility indices in continuous time Markov chains", *Econometrica*, **54**: 1407 – 1423.

[8] Girling, A. J. (2000), "Rank statistics expressible as integrals under P-P-plots and receiver operating characteristics curves", *Journal of the Royal Statistical Society*, **B 62**: 367 – 382.

[9] Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel (1986), *Robust statistics; the approach based on influence functions*, New York: Wiley.

[10] Hillman, A. L. (1980), "Observations on the relation between 'revealed comparative advantage' and comparative advantage as indicated by pre-trade relative prices", *Weltwirtschaftliches Archiv*, **116**: 315 – 321.

[11] Holmgren, E. B. (1995), "The PP-plot as a method of comparing treatment effects", *Journal of the American Statistical Society*, **90**: 360 – 365.

[12] Mushkudiani, N. (2000), *Statistical applications of generalized quantiles: nonparametric tolerance regions and P-P plots*, Ph. D. dissertation, Eindhoven University of Technology.

[13] Proudman, J. and S. Redding (2000), "Evolving patterns of international trade", *Review of International Economics*, **8**(3): 373 – 396.

[14] Wagenvoort, R.and R. Waldman (2002), "On B-robust instrumental variable estimation of the linear model with panel data", *Journal of Econometrics*, **106**(2): 297 – 324.

[15] White, H., (1980), "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity", *Econometrica*, **48**(4): 817 – 838.

[16] Wilk, M. B. and R. Gnanadesikan (1968), "Probability plotting methods for the analysis of data", *Biometrika*, **55**: 1 – 17.

A   Proofs

A.1   Proof of Lemma 1

Consider $\left(t_{1i}^X, t_{2i}^X\right)$. Under Assumption *A1* the n and m realizations of X_n and Y_m are unique, dividing the unit interval into n and m steps of size $1/n$ and $1/m$, respectively. Consequently, $G^{-1}(i/n) = x_i$ for $i = 1, .., n$. To determine the number of observations in Y_m smaller than or equal to x_i consider the joint sample $\{X_n, Y_m\}$. Under Assumption *A2* $R(x_i)$ denotes the unique rank from the joint, ordered sample $\{X_n, Y_m\}$, of which $i-1$ observations relate to X_n and the remainder to Y_m. Accordingly, $n+m-[R(x_i)+i-1]$ observations in Y_m are smaller than x_i, such that $F(G^{-1}(i/n)) = n + m + 1 - i - R(x_i)$. Similarly for (t_{1j}^Y, t_{2j}^Y).

25
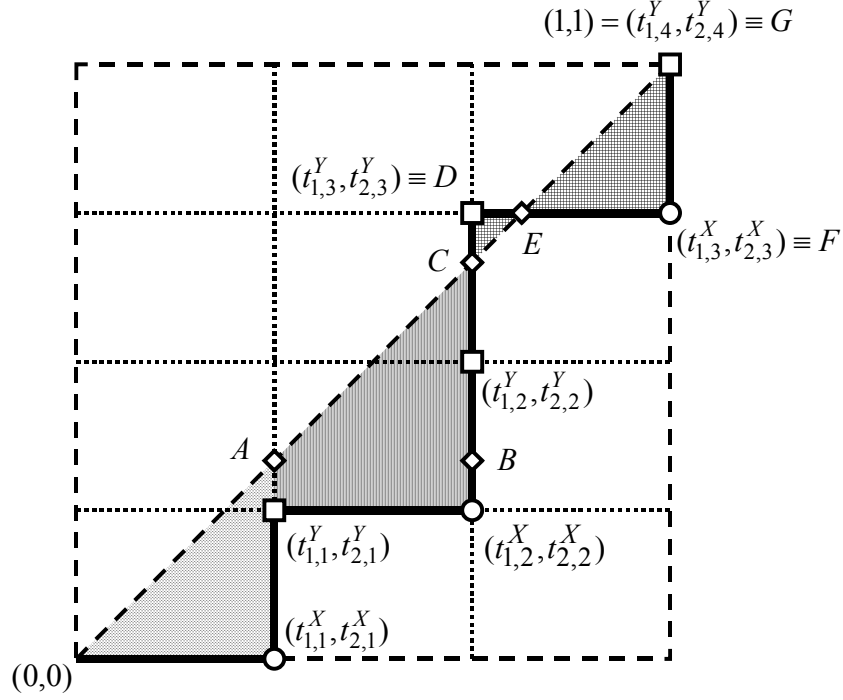
Figure 7: Computation of the sample HM index in case there are no within-sample and no between-sample ties.

A.2 Proof of Lemma 2

Under *A1–A2* the coordinates of the points generated by samples X_n and Y_m are unique and given in Lemma 1, such that an asymmetric discrete sample PP-plot always consists of horizontal and vertical lines only with $t_{1,i}^X, t_{1,j}^Y \in \left\{0, \frac{1}{n}, \frac{2}{n}, ..., 1\right\}$ and $t_{2,i}^X, t_{2,j}^Y \in \left\{0, \frac{1}{m}, \frac{2}{m}, ..., 1\right\}$. If suffices to calculate the sample HM index by summation of the step-wise deviation from the diagonal of the points $\left(t_{1,i}^X, t_{2,i}^X\right)$ generated by sample X_n for $i = 1, .., n$. There are two logical possibilities at step i: (i) the sample PP-plot from $\left(t_{1,i-1}^X, t_{2,i-1}^X\right)$ to $\left(t_{1,i}^X, t_{2,i}^X\right)$ does not cut the diagonal horizontally ($I_i = 0$) or (ii) the sample PP-plot from $\left(t_{1,i-1}^X, t_{2,i-1}^X\right)$ to $\left(t_{1,i}^X, t_{2,i}^X\right)$ does cut the diagonal horizontally ($I_i = 1$). These possibilities are illustrated in Figure 7 (where $n = 3$ and $m = 4$) from $\left(t_{1,1}^X, t_{2,1}^X\right)$ to $\left(t_{1,2}^X, t_{2,2}^X\right)$, and from $\left(t_{1,2}^X, t_{2,2}^X\right)$ to $\left(t_{1,3}^X, t_{2,3}^X\right)$, respectively, where the circles indicate the points generated by X_n, the squares those generated by Y_m, and where the diamonds are auxiliary points. Note that possibility (ii) does not arise for symmetric samples (see also Lemma 3).

Ad (i). If $I_i = 0$, the diagonal is not cut horizontally from $i - 1$ to i,

as illustrated at $i = 2$ in Figure 7. The contribution to the sample HM index from such a step consists of twice the area of the triangle ABC plus twice the area of the rectangle between the sample PP-plot and the points AB (both in case the sample PP-plot would be above or below the diagonal). The former contribution is equal to $\left(t_{1,i}^X - t_{1,i-1}^X\right)^2$ and the latter equals $2\left(t_{1,i}^X - t_{1,i-1}^X\right)\left|t_{2,i}^X - t_{1,i-1}^X\right|$. Since $\left(t_{1,i}^X - t_{1,i-1}^X\right) \geq 0$, their sum simplifies to $\left(t_{1,i}^X - t_{1,i-1}^X\right)\left|2t_{2,i}^X - t_{1,i-1}^X - t_{1,i-1}^X\right|$. Note that the area of the rectangle disappears if the sample PP-plot coincides with the diagonal at $i - 1$, as is the case for $i = 1$ in Figure 7.

Ad (ii). If $I_i = 1$, the diagonal is cut horizontally from $i - 1$ to i, as illustrated at $i = 3$ in Figure 7. The contribution to the sample HM index from such a step consists of twice the area of the triangle CDE plus twice the area of the triangle EFG. As the points C, E, and G are on the diagonal, their coordinates are given by $(t_{1,i-1}^X, t_{1,i-1}^X)$, $(t_{2,i}^X, t_{2,i}^X)$, and $(t_{1,i}^X, t_{1,i}^X)$, respectively. Twice the area of the triangle CDE is therefore equal to $\left(t_{2,i}^X - t_{1,i-1}^X\right)^2$ and twice the area of triangle EFG equals $\left(t_{1,i}^X - t_{2,i}^X\right)^2$.

A.3   Proof of Lemma 3

The proof evolves along two steps. First note that when there are no within-sample and no between-sample ties, a discrete sample PP-plot always consists of horizontal and vertical lines only while both the x- and y-coordinates include all entries in $C \equiv \left\{0, \frac{1}{n}, \frac{2}{n}, ..., 1\right\}$. The absence of between-sample ties precludes diagonal lines, the absence of within sample ties implies all entries in C to be included once as either coordinate. Second, there is a one-to-one correspondence between the absolute difference between the x- and y-coordinates of any point (t_1, t_2) of a discrete sample PP-plot and its contribution to the sample HM index (see e.g. Figure 2); any step of size $\frac{1}{n}$ (either vertical or horizontal) away from the diagonal accounts for a triangle with surface $\frac{1}{2n^2}$. Accordingly, each step increases the sample HM index by $1/n^2$. The contribution of (t_1, t_2) to the sample HM index thus equals: $|t_1 - t_2|/n$ (note that the coordinates are a fraction of n). Summing up over all coordinates then yields the lemma.

A.4   Proof of Proposition 1

Under assumptions $A1 - A3$ the pdf of the sample HM index for symmetric samples of size n, $h\left(HM_n^S\right)$, can be constructed on the basis of the pdf for sample size of $n - 1$, $h\left(HM_{n-1}^S\right)$. In what follows, starting from $n = 1$, this iterative procedure is explained up to $n = 3$. For larger sample sizes the

same logic applies.

First note that sample PP-plots based on $\widetilde{F}_n$ and $\widetilde{G}_n$ can be visualized by dividing the unit square into n^2 blocks of size $1/n^2$. Any possible sample PP-plot is then a non-decreasing path along the edges of these blocks from $(0,0)$ to $(1,1)$. In Figure 8 this is illustrated for $n = 1, 2, 3$.

For $n = 1$ there are two, equally likely pointwise sample comparisons (with concomitant distinct sample PP-plots), both yielding a sample HM index value of one:

$$\begin{aligned} X_1(1) &> Y_1(1); \\ X_1(1) &< Y_1(1). \end{aligned}$$

In the first case the sample PP-plot follows path $O_1 D_{11} F$ in Panel a of Figure 8 while in the second case the sample PP-plot coincides with $O_1 E_{11} F$. Without loss of generality assume that $X_1(n) > Y_1(n)$ (symmetry implies that the actual number of possible sample PP-plots is twice as large without this assumption; it does not, however, affect the number of possible distinct sample HM index values, nor the probability of their occurrence). Using the notation of the main text note:

$$\begin{aligned} \Omega(1) &= 1;\ \Theta(1) = 1;\ HM_1^S = 1;\ M(1) = 1; \\ D(1) &= 1;\ B(1) = \begin{pmatrix} 1 & 2 \end{pmatrix};\ A(2) = \begin{pmatrix} 1 & 2 \end{pmatrix}. \end{aligned}$$

Matrices A, B and M are discussed below.

For $n = 2$, given that $X_1(2) > Y_1(2)$, there are three, equally likely pointwise sample comparisons:

$$\begin{aligned} X_1(2) &> X_2(2) > Y_1(2) > Y_2(2); \\ X_1(2) &> Y_1(2) > X_2(2) > Y_2(2); \\ X_1(2) &> Y_1(2) > Y_2(2) > X_2(2). \end{aligned}$$

In Panel b of Figure 8 these possibilities respectively corresponds to sample PP-plot $O_2 D_{21} D_{22} D_{11} F$, $O_2 D_{21} O_1 D_{11} F$, and $O_2 D_{21} O_1 E_{11} F$. Accordingly, HM_2^S equals 1 with probability $\frac{1}{3}$, and $\frac{1}{2}$ with probability $\frac{2}{3}$.

Observe the iterative structure going from $n = 1$ to $n = 2$. First, considering all possible sample PP-plots from O_1 onwards in Panel b corresponds to examining all possible sample PP-plots as if $n = 1$. The latter are characterized above assuming that $X_1(1) > Y_1(1)$. Since this holds by definition at
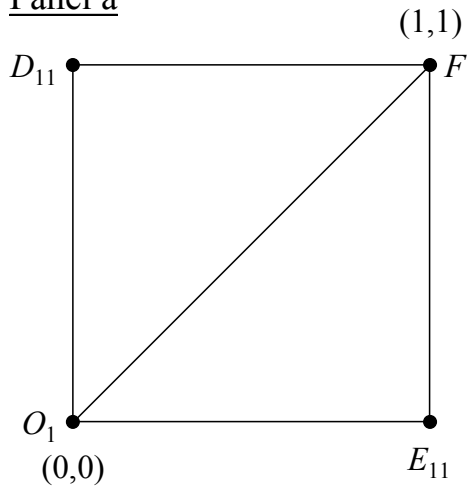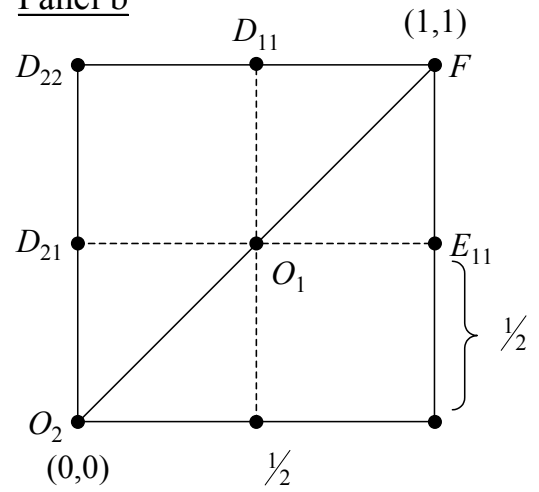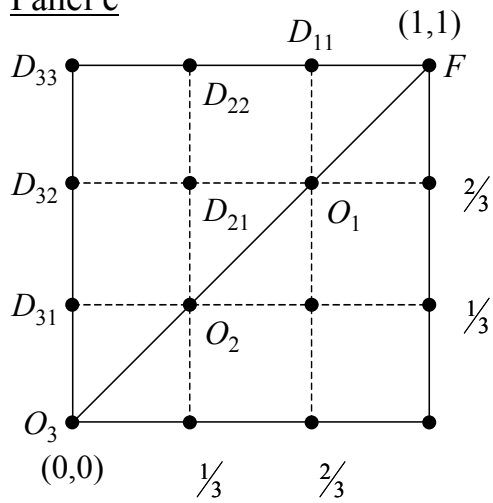
Figure 8: Grid for all possible sample PP-plots for $n = 1$ (Panel a), $n = 2$ (Panel b), and $n = 3$ (Panel c).

O_1 in Figure 8 Panel b, all possible sample PP-plots have to be considered from O_1 onwards. This explains the $2\Upsilon_n$ in matrix $B(n)$ (and in particular the entry "2" in $B(1)$). Second, for all sample PP-plots going through O_1 in Panel b of Figure 8, two times $1/n^2$ – the surface value of block $D_{21}D_{22}D_{11}O_1$ multiplied by two – has to be substracted from one for determining the value of the sample HM index, in addition to all the blocks that have to be substracted in relation to arriving at O_1 in the first place. Third, note that sample PP-plot $O_2D_{21}D_{22}D_{11}F$ can only be reached if $X_1(1) > Y_1(1)$. This explains the first column of matrix $B(n)$ (and in particular the entry "1" in $B(1)$). For $n = 2$ the various components of $h\left(HM_n^S\right)$ read as:

$$\Omega(2) = 3; \ \Theta(2) = 2; \ HM_2^S = \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix}; \ M(2) = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix};$$

$$D(2) = \begin{pmatrix} 1/3 \\ 2/3 \end{pmatrix}; \ B(2) = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 2 \end{pmatrix}; \ A(3) = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 2 & 4 \end{pmatrix}.$$

Matrix $M(n)$ is constructed to summarize all possible sample PP-plots, where the row coordinate – from top to bottom – refers to the number of blocks that have to be substracted, where the column coordinate – from left to right – corresponds to the position – from top to bottom – along the vertical line at location $1/n$ in sample PP-plot space for samples of size n (for example locations D_{11} and O_1 in Figure 8, Panel b), and the matrix entry refers to the number of possible sample PP-plots with the related value of HM_n^S. Hence, the sum of the kth row of $M(n)$ is the number of distinct sample PP-plots for which $k - 1$ times the number of blocks of size $1/n^2$ (times two) has to be subtracted from one in order to determine the value of the sample HM index for these plots (thus being: $1 - 2(k-1)/n^2$). Note that for $n = 1$ zero blocks are subtracted.

Indeed, the rows of $M(2)$ are associated with the number of possible distinct sample PP-plots yielding HM_2^S. Entry "1" indicates that there is one sample PP-plot yielding $HM_2^S(i) = 1$, being $O_2D_{21}D_{22}D_{11}F$, while entry "2" implies that $HM_2^S(i) = \frac{1}{2}$ is obtained for two, distinct sample PP-plots, namely $O_2D_{21}O_1D_{11}F$ and $O_2D_{21}O_1E_{11}F$. The probabilities, HM_2^S, thus follow from summing the rows of $B(2)$ and dividing these by the total number of distinct sample PP-plots, $\Omega(2)$. Auxiliary matrices $B(1)$ and $A(2)$ are defined to construct $M(2)$ on the basis of $M(1)$.

For $n = 3$, given that $X_1(3) > Y_1(3)$, there are ten, equally likewise pointwise sample comparisons giving rise to the various distinct sample PP-plots (Figure 8, Panel c). First note that the upper North-East block starting at O_2 corresponds to the situation for $n = 2$, referred to as the "n_2-block".

However, additional sample PP-plots have to be considered; those that "enter" the n_2-block at either D_{21} or D_{22}. Second, because the $n = 2$ case was characterized assuming $X_1(2) > Y_1(2)$ all possible sample PP-plots from O_2 onwards have to be multiplied by 2 as $X_1(2) < Y_1(2)$ can now also apply (indeed, the $2\Upsilon_n$ again). Note that this does not apply to sample PP-plots "entering" the n_2-block at either D_{21} or D_{22} because for the sample PP-plots going through these points in the $n = 2$ case the assumption $X_1(2) > Y_1(2)$ applies by definition. Third, for determining the value of $HM_3^S(i)$, twice the surface area of block $D_{31}D_{33}D_{22}O_2$ has to be substracted from one for all sample PP-plots "entering" the n_2-block at O_2 (again, in addition to what is to be substracted for arriving at O_2 as such), once the surface area of block $D_{32}D_{33}D_{22}D_{21}$ for all sample PP-plots "entering" the n_2-block at D_{21}, and nothing for all sample PP-plots "entering" the n_2-block at D_{22}. Note that in this case:

$$\Omega(3) = 10; \ \Theta(3) = 4; \ HM_3^S = \begin{pmatrix} 1 \\ 7/9 \\ 5/9 \\ 3/9 \end{pmatrix}; \ M(3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 4 \end{pmatrix};$$

$$D(3) = \begin{pmatrix} 1/10 \\ 1/10 \\ 4/10 \\ 4/10 \end{pmatrix}; \ B(3) = \begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 2 \end{pmatrix}; \ A(4) = \begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & 1 & 1 & 2 \\ 0 & 2 & 4 & 8 \\ 0 & 0 & 4 & 8 \end{pmatrix}.$$

Again, the rows of matrix $M(3)$ are associated with the number of possible distinct sample PP-plots resulting in HM_3^S: $M_{11}(3) = 1$ indicates that there is one possibility left for sample PP-plots to arrive at $(1,1)$ from D_{22} onwards yielding $HM_3^S(1) = 1$, $M_{22}(3) = 1$ implies that starting at D_{21} there is one possibility left for sample PP-plots to reach $(1,1)$ giving $HM_3^S(2) = \frac{7}{9}$, $M_{32}(3) = 2$ signifies there being two possible PP-plots to end at $(1,1)$ from entering the n_2-block at D_{22} resulting in $HM_3^S(3) = \frac{5}{9}$, $M_{33}(3) = 2$ means that two possible distinct sample PP-plots exist that run through O_2 and yield $HM_3^S(3) = \frac{5}{9}$, and $M_{34}(3) = 4$ says that there are four distinct sample PP-plots going through O_2 corresponding to $HM_3^S(3) = \frac{3}{9}$. Summing up the rows of $M(3)$ and dividing these sums by $\Omega(3)$ returns D_3. To construct $M(3)$ on the basis of $M(2)$ the auxiliary matrices $B(2)$ and $A(3)$ follow.

B   The sample HM index CDF

n	z_{90}	z_{95}	$z_{97.5}$	z_{99}	μ_1	μ_2	μ_3	μ_4	σ^2
2	1.0000	1.0000	1.0000	1.0000	0.6667	0.5000	0.4167	0.3750	0.0556
3	0.7778	1.0000	1.0000	1.0000	0.5333	0.3284	0.2305	0.1796	0.0440
4	0.7500	0.8750	1.0000	1.0000	0.4571	0.2437	0.1498	0.1036	0.0348
5	0.6800	0.7600	0.8400	0.9200	0.3694	0.1936	0.1070	0.0670	0.0285
6	0.6111	0.6667	0.7778	0.8333	0.3694	0.1605	0.0812	0.0468	0.0240
7	0.5510	0.6327	0.7143	0.7959	0.3410	0.1370	0.0643	0.0345	0.0208
8	0.5000	0.5938	0.6563	0.7500	0.3183	0.1195	0.0525	0.0264	0.0182
9	0.4815	0.5556	0.6296	0.7037	0.2995	0.1060	0.0439	0.0209	0.0163
10	0.4600	0.5200	0.5800	0.6600	0.2838	0.0952	0.0374	0.0169	0.0147
11	0.4380	0.5041	0.5537	0.6364	0.2703	0.0864	0.0324	0.0140	0.0134
12	0.4167	0.4722	0.5417	0.6111	0.2585	0.0791	0.0284	0.0118	0.0123
13	0.3964	0.4556	0.5148	0.5858	0.2482	0.0729	0.0252	0.0100	0.0113
14	0.3776	0.4388	0.5000	0.5612	0.2390	0.0676	0.0225	0.0086	0.0105
15	0.3689	0.4311	0.4844	0.5467	0.2307	0.0631	0.0203	0.0075	0.0098
16	0.3594	0.4141	0.4688	0.5313	0.2233	0.0591	0.0184	0.0066	0.0092
17	0.3426	0.3979	0.4533	0.5156	0.2165	0.0556	0.0168	0.0059	0.0087
18	0.3333	0.3889	0.4383	0.5000	0.2103	0.0524	0.0154	0.0052	0.0082
19	0.3241	0.3795	0.4294	0.4848	0.2047	0.0497	0.0142	0.0047	0.0078
20	0.3200	0.3700	0.4150	0.4750	0.1994	0.0472	0.0131	0.0042	0.0074
21	0.3107	0.3605	0.4059	0.4603	0.1945	0.0449	0.0122	0.0038	0.0070
22	0.3017	0.3512	0.3967	0.4504	0.1900	0.0428	0.0114	0.0035	0.0067
23	0.2968	0.3459	0.3875	0.4405	0.1858	0.0409	0.0106	0.0032	0.0064
24	0.2882	0.3368	0.3819	0.4340	0.1818	0.0392	0.0100	0.0029	0.0062
25	0.2832	0.3312	0.3728	0.4240	0.1781	0.0376	0.0094	0.0027	0.0059
26	0.2781	0.3225	0.3669	0.4142	0.1746	0.0362	0.0088	0.0025	0.0057
27	0.2730	0.3169	0.3580	0.4074	0.1713	0.0348	0.0084	0.0023	0.0055
28	0.2679	0.3112	0.3520	0.4005	0.1682	0.0336	0.0079	0.0022	0.0053
29	0.2628	0.3056	0.3460	0.3936	0.1653	0.0324	0.0075	0.0020	0.0051
30	0.2600	0.3022	0.3400	0.3867	0.1625	0.0313	0.0071	0.0019	0.0049
31	0.2549	0.2966	0.3340	0.3798	0.1598	0.0303	0.0068	0.0018	0.0048
32	0.2500	0.2910	0.3301	0.3750	0.1573	0.0294	0.0065	0.0017	0.0046
33	0.2470	0.2874	0.3242	0.3701	0.1549	0.0285	0.0062	0.0016	0.0045

Table 8: Critical percentiles, first four moments, and the variance of the sample HM index CDF under H0 for symmetric samples.

n	z_{90}	z_{95}	$z_{97.5}$	z_{99}	μ_1	μ_2	μ_3	μ_4	σ^2
34	0.2439	0.2837	0.3201	0.3633	0.1525	0.0276	0.0059	0.0015	0.0044
35	0.2392	0.2784	0.3159	0.3584	0.1503	0.0268	0.0057	0.0014	0.0042
36	0.2361	0.2747	0.3102	0.3534	0.1482	0.0261	0.0054	0.0013	0.0041
37	0.2330	0.2710	0.3061	0.3484	0.1462	0.00254	0.0052	0.0012	0.0040
38	0.2299	0.2673	0.3019	0.3435	0.1442	0.0247	0.0050	0.0012	0.0039
39	0.2268	0.2636	0.2991	0.3399	0.1424	0.0241	0.0048	0.0011	0.0038
40	0.2238	0.2612	0.2950	0.3350	0.1406	0.0235	0.0046	0.0011	0.0037
41	0.2219	0.2576	0.2909	0.3314	0.1388	0.0229	0.0045	0.0010	0.0036
42	0.2188	0.2551	0.2880	0.3277	0.1372	0.0223	0.0043	0.0010	0.0035
43	0.2158	0.2515	0.2839	0.3240	0.1355	0.0218	0.0041	0.0009	0.0034
44	0.2138	0.2490	0.2810	0.3202	0.1340	0.0213	0.0040	0.0009	0.0034
45	0.2109	0.2454	0.2780	0.3165	0.1325	0.0208	0.0039	0.0008	0.0033
46	0.2089	0.2429	0.2750	0.3129	0.1310	0.0204	0.0037	0.0008	0.0032
47	0.2069	0.2404	0.2721	0.3092	0.1296	0.0199	0.0036	0.0008	0.0031
48	0.2049	0.2378	0.2691	0.3064	0.1282	0.0195	0.0035	0.0007	0.0031
49	0.2020	0.2353	0.2661	0.3036	0.1269	0.0191	0.0034	0.0007	0.0030
50	0.2000	0.2336	0.2640	0.3000	0.1256	0.0187	0.0033	0.0007	0.0030
51	0.1980	0.2311	0.2611	0.2972	0.1244	0.0184	0.0032	0.0007	0.0029
52	0.1960	0.2286	0.2581	0.2944	0.1232	0.0180	0.0031	0.0006	0.0028
53	0.1947	0.2268	0.2560	0.2916	0.1220	0.0177	0.0030	0.0006	0.0028
54	0.1927	0.2243	0.2538	0.2888	0.1209	0.0174	0.0029	0.0006	0.0027
55	0.1907	0.2225	0.2516	0.2866	0.1198	0.0170	0.0029	0.0006	0.0027
56	0.1894	0.2207	0.2494	0.2838	0.1187	0.0167	0.0028	0.0005	0.0026
57	0.1874	0.2182	0.2472	0.2810	0.1176	0.0164	0.0027	0.0005	0.0026
58	0.1861	0.2164	0.2449	0.2788	0.1166	0.0162	0.0026	0.0005	0.0026
59	0.1841	0.2146	0.2427	0.2766	0.1156	0.0159	0.0026	0.0005	0.0025
60	0.1828	0.2128	0.2406	0.2739	0.1147	0.0156	0.0025	0.0005	0.0025
61	0.1814	0.2110	0.2389	0.2717	0.1137	0.0154	0.0025	0.0005	0.0024
62	0.1800	0.2092	0.2367	0.2695	0.1128	0.0151	0.0024	0.0004	0.0024
63	0.1781	0.2079	0.2351	0.2673	0.1119	0.0149	0.0023	0.0004	0.0023
64	0.1768	0.2061	0.2329	0.2656	0.1110	0.0146	0.0023	0.0004	0.0023
65	0.1754	0.2043	0.2312	0.2634	0.1101	0.0144	0.0022	0.0004	0.0023
66	0.1740	02029	0.2296	0.2612	0.1093	0.0142	0.0022	0.0004	0.0022
67	0.1731	0.2016	0.2279	0.2595	0.1085	0.0140	0.0021	0.0004	0.0022

Table 9: Continued

n	z_{90}	z_{95}	$z_{97.5}$	z_{99}	μ_1	μ_2	μ_3	μ_4	σ^2
68	0.1717	0.1998	0.2262	0.2574	0.1077	0.0138	0.0021	0.0004	0.0022
69	0.1703	0.1985	0.2245	0.2556	0.1069	0.0136	0.0020	0.0004	0.0021
70	0.1690	0.1971	0.2229	0.2539	0.1061	0.0134	0.0020	0.0003	0.0021
71	0.1680	0.1958	0.2212	0.2521	0.1054	0.0132	0.0020	0.0003	0.0021
72	0.1667	0.1944	0.2195	0.2504	0.1046	0.0130	0.0019	0.0003	0.0021
73	0.1657	0.1931	0.2182	0.2486	0.1039	0.0128	0.0019	0.0003	0.0020
74	0.1644	0.1917	0.2166	0.2469	0.1032	0.0126	0.0018	0.0003	0.0020
75	0.1634	0.1904	0.2153	0.2452	0.1025	0.0125	0.0018	0.0003	0.0020
76	0.1624	0.1891	0.2136	0.2438	0.1018	0.0123	0.0018	0.0003	0.0019
77	0.1611	0.1877	0.2123	0.2420	0.1012	0.0122	0.0017	0.0003	0.0019
78	0.1601	0.1867	0.2110	0.2406	0.1005	0.0120	0.0017	0.0003	0.0019
79	0.1591	0.1854	0.2097	0.2389	0.0999	0.0118	0.0017	0.0003	0.0019
80	0.1581	0.1844	0.2084	0.2375	0.0992	0.0117	0.0016	0.0003	0.0018
81	0.1571	0.1831	0.2071	0.2361	0.0986	0.0116	0.0016	0.0003	0.0018
82	0.1562	0.1820	0.2058	0.2347	0.0980	0.0114	0.0016	0.0003	0.0018
83	0.1552	0.1810	0.2045	0.2333	0.0974	0.0113	0.0015	0.0002	0.0018
84	0.1545	0.1797	0.2032	0.2319	0.0968	0.0111	0.0015	0.0002	0.0018
85	0.1535	0.1787	0.2022	0.2304	0.0963	0.0110	0.0015	0.0002	0.0017
86	0.1525	0.1777	0.2009	0.2290	0.0957	0.0109	0.0015	0.0002	0.0017
87	0.1515	0.1766	0.1999	0.2279	0.0952	0.0108	0.0014	0.0002	0.0017
88	0.1508	0.1756	0.1986	0.2265	0.0946	0.0106	0.0014	0.0002	0.0017
89	0.1499	0.1746	0.1976	0.2251	0.0941	0.0105	0.0014	0.0002	0.0017
90	0.1491	0.1738	0.1965	0.2240	0.0935	0.0104	0.0014	0.0002	0.0016
91	0.1482	0.1728	0.1953	0.2228	0.0930	0.0103	0.0013	0.0002	0.0016
92	0.1474	0.1718	0.1942	0.2214	0.0925	0.0102	0.0013	0.0002	0.0016
93	0.1467	0.1710	0.1932	0.2203	0.0920	0.0101	0.0013	0.0002	0.0016
94	0.1460	0.1700	0.1922	0.2191	0.0915	0.0100	0.0013	0.0002	0.0016
95	0.1450	0.1690	0.1911	0.2180	0.0910	0.0098	0.0013	0.0002	0.0016
96	0.1443	0.1682	0.1903	0.2168	0.0906	0.0097	0.0012	0.0002	0.0015
97	0.1436	0.1674	0.1893	0.2156	0.0901	0.0096	0.0012	0.0002	0.0015
98	0.1429	0.1664	0.1883	0.2147	0.0896	0.0095	0.0012	0.0002	0.0015
99	0.1421	0.1656	0.1872	0.2135	0.0892	0.0094	0.0012	0.0002	0.0015
100	0.1414	0.1648	0.1864	0.2124	0.0887	0.0094	0.0012	0.0002	0.0015
101	0.1407	0.1640	0.1854	0.2114	0.0883	0.0093	0.0011	0.0002	0.0015

Table 10: Continued

n	z_{90}	z_{95}	$z_{97.5}$	z_{99}	μ_1	μ_2	μ_3	μ_4	σ^2
102	0.1399	0.1632	0.1845	0.2103	0.0879	0.0092	0.0011	0.0002	0.0015
103	0.1394	0.1624	0.1837	0.2094	0.0574	0.0091	0.0011	0.0002	0.0014
104	0.1387	0.1616	0.1827	0.2084	0.0870	0.0090	0.0011	0.0002	0.0014
105	0.1380	0.1608	0.1819	0.2074	0.0866	0.0089	0.0011	0.0002	0.0014
106	0.1374	0.1600	0.1810	0.2065	0.0862	0.0088	0.0011	0.0002	0.0014
107	0.1367	0.1592	0.1802	0.2053	0.0858	0.0087	0.0011	0.0001	0.0014
108	0.1361	0.1586	0.1794	0.2046	0.0854	0.0087	0.0010	0.0001	0.0014
109	0.1354	0.1578	0.1785	0.2036	0.0850	0.0086	0.0010	0.0001	0.0014
110	0.1349	0.1572	0.1777	0.2026	0.0846	0.0085	0.0010	0.0001	0.0013
111	0.1342	0.1564	0.1769	0.2017	0.0842	0.0084	0.0010	0.0001	0.0013
112	0.1336	0.1558	0.1762	0.2007	0.0838	0.0083	0.0010	0.0001	0.0013
113	0.1331	0.1550	0.1753	0.1999	0.0835	0.0083	0.0010	0.0001	0.0013
114	0.1325	0.1544	0.1745	0.1990	0.0831	0.0082	0.0010	0.0001	0.0013
115	0.1318	0.1537	0.1738	0.1982	0.0827	0.0081	0.0009	0.0001	0.0013
116	0.1312	0.1529	0.1730	0.1974	0.0824	0.0081	0.0009	0.0001	0.0013
117	0.1307	0.1523	0.1723	0.1964	0.0820	0.0080	0.0009	0.0001	0.0013
118	0.1301	0.1517	0.1715	0.1956	0.0817	0.0079	0.0009	0.0001	0.0013
119	0.1296	0.1510	0.1708	0.1948	0.0813	0.0079	0.0009	0.0001	0.0012
120	0.1290	0.1504	0.1701	0.1940	0.0810	0.0078	0.0009	0.0001	0.0012
121	0.1286	0.1498	0.1695	0.1932	0.0806	0.0077	0.0009	0.0001	0.0012
122	0.1281	0.1492	0.1688	0.1924	0.0803	0.0077	0.0009	0.0001	0.0012
123	0.1275	0.1485	0.1681	0.1916	0.0800	0.0076	0.0009	0.0001	0.0012
124	0.1270	0.1480	0.1674	0.1908	0.0797	0.0075	0.0008	0.0001	0.0012
125	0.1265	0.1474	0.1667	0.1901	0.0793	0.0075	0.0008	0.0001	0.0012
126	0.1260	0.1468	0.1660	0.1893	0.0790	0.0074	0.0008	0.0001	0.0012
127	0.1254	0.1463	0.1654	0.1885	0.0787	0.0074	0.0008	0.0001	0.0012
128	0.1250	0.1456	0.1647	0.1879	0.0784	0.0073	0.0008	0.0001	0.0012
129	0.1245	0.4151	0.1641	0.1871	0.0781	0.0072	0.0008	0.0001	0.0011
130	0.1240	0.1445	0.1634	0.1864	0.0778	0.0072	0.0008	0.0001	0.0011
131	0.1235	0.1440	0.1629	0.1857	0.0775	0.0071	0.0008	0.0001	0.0011
132	0.1230	0.1434	0.1622	0.1850	0.0772	0.0071	0.0008	0.0001	0.0011
133	0.1226	0.1429	0.1616	0.1842	0.0769	0.0070	0.0008	0.0001	0.0011
134	0.1222	0.1423	0.1609	0.1836	0.0766	0.0070	0.0008	0.0001	0.0011
135	0.1216	0.1418	0.1604	0.1829	0.0763	0.0069	0.0007	0.0001	0.0011

Table 11: Continued

n	z_{90}	z_{95}	$z_{97.5}$	z_{99}	μ_1	μ_2	μ_3	μ_4	σ^2
136	0.1212	0.1413	0.1598	0.1822	0.0761	0.0069	0.0007	0.0001	0.0011
137	0.1208	0.1407	0.1593	0.1816	0.0758	0.0068	0.0007	0.0001	0.0011
138	0.1204	0.1402	0.1587	0.1809	0.0755	0.0068	0.0007	0.0001	0.0011
139	0.1199	0.1397	0.1581	0.1803	0.0752	0.0067	0.0007	0.0001	0.0011
140	0.1195	0.1393	0.1576	0.1796	0.0750	0.0067	0.0007	0.0001	0.0011
141	0.1191	0.1388	0.1570	0.1790	0.0747	0.0066	0.0007	0.0001	0.0010
142	0.1186	0.1383	0.1564	0.1783	0.0744	0.0066	0.0007	0.0001	0.0010
143	0.1182	0.1378	0.1559	0.1778	0.0742	0.0065	0.0007	0.0001	0.0010
144	0.1178	0.1372	0.1553	0.1771	0.0739	0.0065	0.0007	0.0001	0.0010
145	0.1174	0.1368	0.1548	0.1765	0.0737	0.0064	0.0007	0.0001	0.0010
146	0.1170	0.1363	0.1543	0.1759	0.0734	0.0064	0.0007	0.0001	0.0010
147	0.1166	0.1359	0.1537	0.1753	0.0732	0.0064	0.0007	0.0001	0.0010
148	0.1162	0.1354	0.1532	0.1747	0.0729	0.0063	0.0006	0.0001	0.0010
149	0.1158	0.1350	0.1527	0.1741	0.0727	0.0063	0.0006	0.0001	0.0010
150	0.1155	0.1345	0.1522	0.1735	0.0724	0.0062	0.0006	0.0001	0.0010
151	0.1150	0.1341	0.1517	0.1729	0.0722	0.0062	0.0006	0.0001	0.0010
152	0.1147	0.1337	0.1511	0.1724	0.0719	0.0061	0.0006	0.0001	0.0010
153	0.1143	0.1332	0.1507	0.1719	0.0717	0.0061	0.0006	0.0001	0.0010
154	0.1139	0.1327	0.1502	0.1713	0.0715	0.0061	0.0006	0.0001	0.0010
155	0.1135	0.1323	0.1497	0.1707	0.0712	0.0060	0.0006	0.0001	0.0010
156	0.1132	0.1319	0.1492	0.1702	0.0710	0.0060	0.0006	0.0001	0.0009
157	0.1128	0.1315	0.1488	0.1696	0.0708	0.0060	0.0006	0.0001	0.0009
158	0.1125	0.1311	0.1483	0.1691	0.0706	0.0059	0.0006	0.0001	0.0009
159	0.1121	0.1307	0.1478	0.1685	0.0703	0.0059	0.0006	0.0001	0.0009
160	0.1117	0.1302	0.1473	0.1680	0.0701	0.0058	0.0006	0.0001	0.0009
161	0.1114	0.1298	0.1469	0.1675	0.0699	0.0058	0.0006	0.0001	0.0009
162	0.1110	0.1294	0.1464	0.1670	0.0697	0.0058	0.0006	0.0001	0.0009
163	0.1107	0.1291	0.1460	0.1665	0.0695	0.0057	0.0006	0.0001	0.0009
164	0.1104	0.1286	0.1455	0.1660	0.0693	0.0057	0.0006	0.0001	0.0009
165	0.1100	0.1282	0.1451	0.1655	0.0690	0.0057	0.0005	0.0001	0.0009
166	0.1097	0.1279	0.1447	0.1650	0.0688	0.0056	0.0005	0.0001	0.0009
167	0.1094	0.1275	0.1442	0.1645	0.0686	0.0056	0.0005	0.0001	0.0009
168	0.1091	0.1271	0.1438	0.1640	0.0684	0.0056	0.0005	0.0001	0.0009
169	0.1087	0.1267	0.1434	0.1635	0.0682	0.0055	0.0005	0.0001	0.0009

Table 12: Continued

n	z_{90}	z_{95}	$z_{97.5}$	z_{99}	μ_1	μ_2	μ_3	μ_4	σ^2
170	0.1084	0.1264	0.1429	0.1630	0.0680	0.0055	0.0005	0.0001	0.0009
171	0.1081	0.1260	0.1425	0.1625	0.0678	0.0055	0.0005	0.0001	0.0009
172	0.1078	0.1256	0.1421	0.1620	0.0676	0.0054	0.0005	0.0001	0.0009
173	0.1075	0.1253	0.1417	0.1616	0.0674	0.0054	0.0005	0.0001	0.0009
174	0.1071	0.1249	0.1413	0.1611	0.0672	0.0054	0.0005	0.0001	0.0009
175	0.1069	0.1245	0.1409	0.1607	0.0670	0.0053	0.0005	0.0001	0.0008
176	0.1065	0.1242	0.1405	0.1603	0.0668	0.0053	0.0005	0.0001	0.0008
177	0.1063	0.1238	0.1401	0.1598	0.0667	0.0053	0.0005	0.0001	0.0008
178	0.1059	0.1235	0.1397	0.1593	0.0665	0.0052	0.0005	0.0001	0.0008
179	0.1056	0.1231	0.1393	0.1589	0.0663	0.0052	0.0005	0.0001	0.0008
180	0.1054	0.1228	0.1389	0.1585	0.0661	0.0052	0.0005	0.0001	0.0008
181	0.1050	0.1224	0.1385	0.1580	0.0659	0.0052	0.0005	0.0001	0.0008
182	0.1048	0.1221	0.1381	0.1576	0.0657	0.0051	0.0005	0.0001	0.0008
183	0.1045	0.1218	0.1377	0.1572	0.0656	0.0051	0.0005	0.0001	0.0008
184	0.1042	0.1215	0.1374	0.1567	0.0654	0.0051	0.0005	0.0000	0.0008
185	0.1039	0.1211	0.1370	0.1563	0.0652	0.0051	0.0005	0.0000	0.0008
186	0.1037	0.1208	0.1367	0.1559	0.0650	0.0050	0.0005	0.0000	0.0008
187	0.1034	0.1205	0.1363	0.1554	0.0649	0.0050	0.0005	0.0000	0.0008
188	0.1031	0.1201	0.1359	0.1550	0.0647	0.0050	0.0005	0.0000	0.0008
189	0.1028	0.1198	0.1356	0.1546	0.0645	0.0049	0.0004	0.0000	0.0008
190	0.1025	0.1195	0.1352	0.1542	0.0643	0.0049	0.0004	0.0000	0.0008
191	0.1023	0.1192	0.1348	0.1538	0.0642	0.0049	0.0004	0.0000	0.0008
192	0.1020	0.1189	0.1345	0.1534	0.0640	0.0049	0.0004	0.0000	0.0008
193	0.1017	0.1186	0.1342	0.1530	0.0638	0.0048	0.0004	0.0000	0.0008
194	0.1015	0.1183	0.1338	0.1526	0.0637	0.0048	0.0004	0.0000	0.0008
195	0.1012	0.1179	0.1335	0.1522	0.0635	0.0048	0.0004	0.0000	0.0008
196	0.1009	0.1177	0.1331	0.1519	0.0633	0.0048	0.0004	0.0000	0.0008
197	0.1007	0.1174	0.1328	0.1514	0.0632	0.0047	0.0004	0.0000	0.0008
198	0.1004	0.1171	0.1324	0.1511	0.0630	0.0047	0.0004	0.0000	0.0007
199	0.1002	0.1168	0.1321	0.1507	0.0629	0.0047	0.0004	0.0000	0.0007
200	0.1000	0.1165	0.1318	0.1503	0.0627	0.0047	0.0004	0.0000	0.0007
250	0.0894	0.1042	0.1179	0.1345	0.0561	0.0037	0.0003	0.0000	0.0006
300	0.0816	0.0951	0.1076	0.1228	0.0512	0.0031	0.0002	0.0000	0.0005
350	0.0755	0.0880	0.0996	0.1136	0.0474	0.0027	0.0002	0.0000	0.0004

Table 13: Continued