A k-sample homogeneity test: the Harmonic Weighted Mass index

Jeroen Hinloopen® **Rien J.L.M. Wagenvoort** **Charles van Marrewijk**

University of Amsterdam European Investment Bank Utrecht University

ABSTRACT

We propose a quantification of the p-p plot that assigns equal weight to all distances between the respective distributions: the surface between the p-p plot and the diagonal. This surface is labeled the Harmonic Weighted Mass (HWM) index. We introduce the diagonal-deviation (d-d) plot that allows the index to be computed exactly under all circumstances. This two-dimensional d-d plot accommodates a straightforward extension to the k-sample HWM index, with $k > 2$. A Monte Carlo simulation based on an example involving long-term sovereign credit ratings illustrates the power of the HWM test.

Key words: *k-sample EDF test, p-p plot, d-d plot, Power*
JEL Classifications: C12, C14

1. INTRODUCTION

To determine whether samples are drawn from the same distribution, Empirical Distribution Function (EDF) tests can be used if the underlying population distributions are not known.[1] EDF tests quantify in one way or the other percentile-percentile (p-p) plots: the scatter plot of percentiles of two distributions for all entries of their joint support.[2] In this paper we introduce a new EDF statistic: the Harmonic Weighted Mass (HWM) index. It corresponds to the surface between the (k-dimensional) p-p plot and the diagonal, up to a scaling factor that depends on sample sizes.

For two balanced samples without ties the HWM index closely resembles the L_1–version of the Fisz-Cramér-von Mises statistic (hereafter denoted by L_1–FCvM, see Schmidt and Trede, 1995). The HWM index differs from the L_1–FCvM statistic when there are ties in that it is

---

®Jeroen Hinloopen, University of Amsterdam, FEB/ASE, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands, (email: J.Hinloopen@uva.nl).
Rien J.L.M. Wagenvoort, European Investment Bank, Economic and Financial Studies, 100 Boulevard Konrad Adenauer, Luxemburg, Luxemburg; (email: wagenvoo@eib.org).
Charles van Marrewijk, Utrecht University, Utrecht School of Economics, Kriekenpitplein 21-22, 3584 EC Utrecht, The Netherlands; (email: J.G.M.vanMarrewijk@uu.nl).

We gratefully acknowledge the comments of the Editor (Asad Zaman) and of two anonymous referees. Thanks are further due to Robert Waldmann and Sanne Zwart, to seminar participants at the University of Adelaide, the University of Amsterdam, the University of Cyprus, and the Erasmus University Rotterdam, and at the CFE-ERCIM 2010 conference (London) for constructive inputs. All computing codes (in GAUSS) are available upon request. The usual disclaimer applies.

[1] Examples include the Kolmogorov-Smirnov (KS) test (see Kolmogorov (1933) and Smirnov, 1939), the Fisz-Cramér-von Mises (FCvM) test (see Cramér (1928), Fisz (1960) and von Mises, 1931), the Kuiper (K) test (see Kuiper, 1960), and the Anderson-Darling (AD) test (see Anderson and Darling, 1952).

[2] The obvious alternative is the scatter plot of order statistics of two distributions, the so-called quantile-quantile (q-q) plot. If distributions differ in scale and location only, q-q plots consist of straight lines, which is an attractive property when only differences in the shape of the respective distributions are of interest (Wilk and Gnanadesikan, 1968). But this is a drawback, of course, if differences in scale and location are to be revealed as well (see also Holmgren, 1995).

invariant to the position of the tie in the sequence of order statistics. This makes it a more robust statistic.

Extending the HWM index to the simultaneous comparison of $k > 2$ samples requires first the introduction of the k-dimensional p-p plot. The diagonal of this plot corresponds to the line that cuts all 2-dimensional spaces in equal halves. The k-sample HWM index then corresponds to the surface between this 'diagonal' and the k-dimensional p-p plot. This surface is not uniquely defined however. An obvious norm is to consider for every point the shortest distance between the k-dimensional p-p plot and the diagonal.

We derive the expression with which the HWM index can be computed exactly under all circumstances. For that we introduce the diagonal-deviation (d-d) plot: the projection of the p–p plot onto the diagonal.

Being able to compute the HWM index exactly allows us to simulate its distribution in all cases for which it is not known. For two balanced samples without ties, Hinloopen and Wagenvoort (2010) derive analytically the distribution of the L_1–FCvM statistic. For this special case that distribution also applies to the HWM index. Our Monte Carlo experiments show that in most cases the effect of ties on the concomitant critical values of the 2-sample HWM index is negligible, and that the sample size correction factor deals adequately with unbalanced samples. For $k > 2$ we provide significance tables up to $k = 15$.

Evidently, none of the existing EDF tests dominates any of the other under all circumstances. Strictly speaking therefore, any sample comparison must involve the computation of all tests to rule out type-II errors as much as possible. Schmidt and Trede (1995) show that the L_1–version of the FCvM statistic has nearly the same power as the classical L_2–version of the FCvM statistic when samples are drawn from a set of widely used continuous distributions: Pareto, Log-normal, and Singh-Maddala. This result extends to the HWM index since, in these cases, ties are absent.

To assess the power of the HWM test, we perform a Monte Carlo simulation based on an example involving sovereign credit ratings. It appears that the extended Anderson-Darling (AD) test (see Scholz and Stephens, 1987) has more power than the HWM test when one of the three main rating distributions (Fitch, Moody's and Standard & Poor) is replaced by a uniform distribution. However, the HWM test can outperform AD if, in addition to the distribution homogeneity assumption, also the independency assumption is violated. Indeed, the HWM test is expected to have more power than any of the other EDF tests when the p-p plot remains 'close' to the diagonal over the entire probability space (while samples are drawn from different distributions) as it is the only EDF test that assigns equal weight to all distances between the respective distributions. It should therefore join the basket of EDF tests, also because in these situations a visual inspection of the p-p plot easily leads to incorrect conclusions.

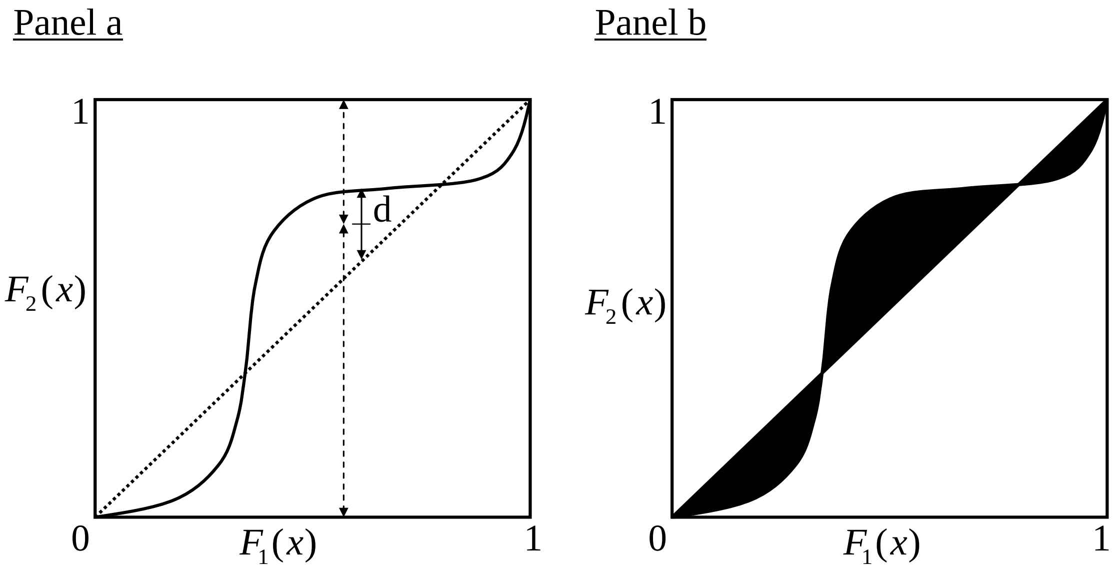
2. THE 2-SAMPLE HWM INDEX

Let Ξ_1 be the set of cumulative distribution functions. The p-p plot based on F_1 and F_2 belonging to Ξ_1 depicts for every domain value z from the joint support of F_1 and F_2 the percentiles of one distribution relative to the other:

$$z \mapsto \begin{bmatrix} F_1(z) \\ F_2(z) \end{bmatrix}. \tag{2.1}$$

This is a plot in the 2-dimensional simplex. An example of a p-p plot for two continuous random variables is given in Panel a of Figure 2.1. Clearly p-p plots are mappings from [0, 1] onto [0, 1] and depict the correspondence between the two underlying distributions in probability space. They are key to the hypothesis underlying EDF tests: H_0: $\hat{F}_1 = \hat{F}_2$, where $\hat{F}$ is an Empirical Distribution Function. For ease of notation, in the remainder of the paper we use F_i, $i = 1, \ldots, K$ to denote an empirical distribution function (rather than a cumulative distribution function).

Figure 2.1 Theoretical p-p plot (panel a) and HWM index (panel b).



2.1. Definition

Because the p-p plot coincides with the diagonal if, and only if, the two underlying distributions are identical, there are various well-known statistics to test H_0 that are based on the distance between the p-p plot and the diagonal (see Hinloopen and Wagenvoort (2010) for an overview). In particular, the L_2–version of the FCvM test sums up over all squared distances d (see Figure 2.1, Panel a) and, graphically, the Anderson-Darling test augments the FCvM-test by weighing every squared distance with the product of the distance between 0 and the centre of d, and the distance between 1 and the centre of d. We propose the area between the diagonal and the p-p plot as the criterion for validating H_0 (see Panel b of Figure 2.1). Because this area reflects the extent to which the probability mass of the two underlying distributions is ‘in harmony’, we label it the Harmonic Weighted Mass (HWM) index.

The HWM index requires a continuous p-p plot. But for discrete random variables the p-p plot is also discrete. The continuous analogue of a discrete p-p plot is obtained by connecting the points of the discrete p-p plot through straight lines (see Girling, 2000). Let X_1 and X_2 be two random variables with empirical distribution functions $F_1(x)$ and $F_2(x)$ respectively. The coordinates of the resulting piece-wise linear continuous p-p plot can be defined as:

$$\begin{pmatrix} \tilde{F}_1(x) \\ \tilde{F}_2(x) \end{pmatrix} \equiv \left\{ \alpha \begin{pmatrix} F_1(z_{i-1}) \\ F_2(z_{i-1}) \end{pmatrix} + (1-\alpha) \begin{pmatrix} F_1(z_i) \\ F_2(z_i) \end{pmatrix} \right\},$$

where $\mathbf{z} \equiv \{z_1, \ldots, z_m\}$ are the ordered domain values of the joint support of X_1 and X_2, with $0 \le \alpha \le 1$.

The 2-sample HWM index is then defined as:

Definition 2.1. $HWM(F_1, F_2) \equiv S(n_1, n_2) \int_{-\infty}^{\infty} \left| \tilde{F}_2(x) - \tilde{F}_1(x) \right| d\tilde{F}_1(x),$

where the sample size correction factor $S(n_1, n_2)$ is taken from Rosenblatt (1952) and Fisz (1960). It reads as:

$$S(n_1, n_2) = \sqrt{\frac{n_1 n_2}{n_1 + n_2}}, \tag{2.2}$$

and speeds up convergence of the finite sample distribution towards the limiting distribution. It also makes the analytical critical percentile values of the HWM index good approximations for unbalanced samples (see Section 2.3).

By virtue of the underlying p-p plot the HWM index is a non-parametric and distribution free mapping. It has two further properties (proofs of all properties are in Appendix A, Section A.1):

Property 2.P1. *(equality):* $HWM(F_1, F_2) = 0 \Leftrightarrow \forall z \in [a, b] : F_1(z) = F_2(z).$

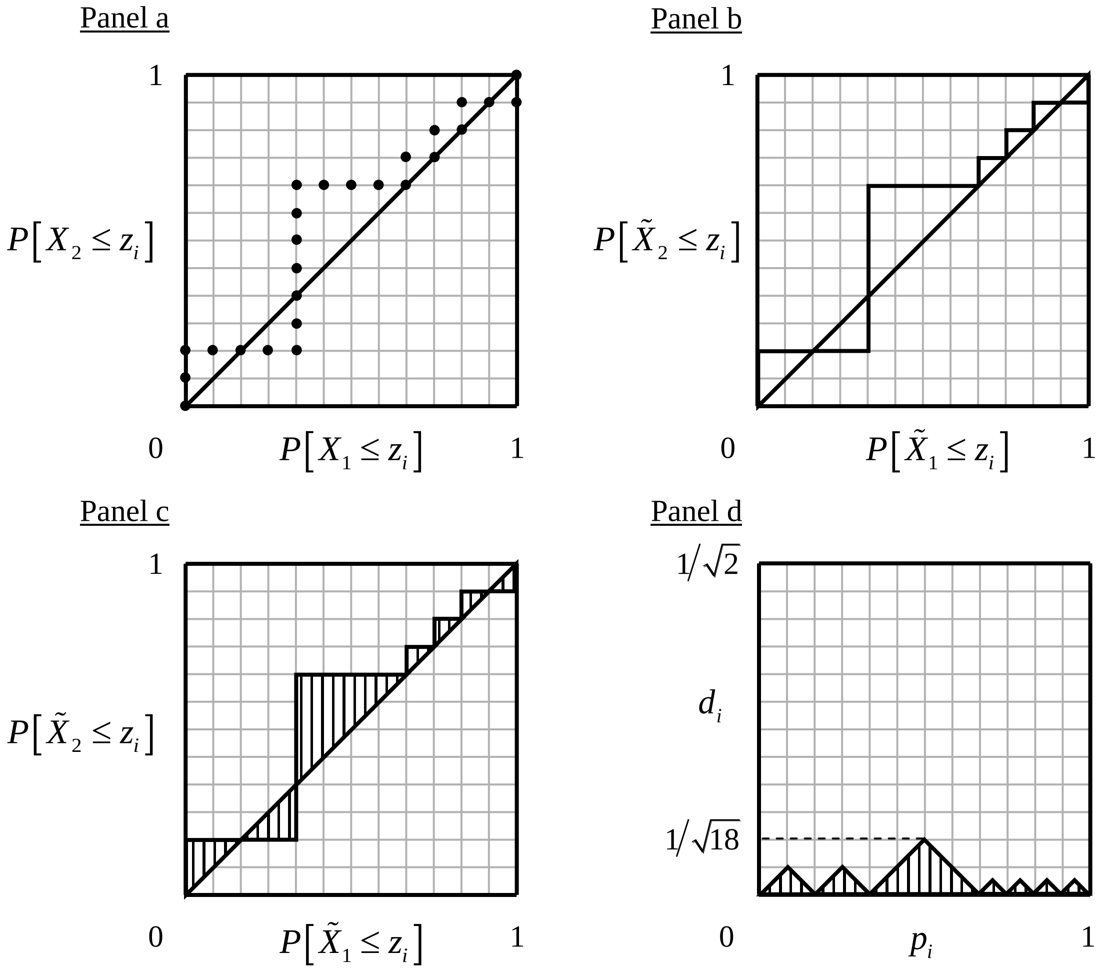
Property 2.P2. *(order irrelevance):* $HWM(F_1, F_2) = HWM(F_2, F_1).$

As an illustration consider the water level of the river Meuse as it enters The Netherlands at Borgharen Dorp in 1990 and 1993 (see Hinloopen, 1997, Chapter 5). In 1993 the Southern part of Holland was plagued by severe floods and it is of interest to know whether the entire year 1993 experienced exceptionally high water levels. Table 2.1 contains the maximum water levels in millimeters for both 1990 and 1993 recorded on each last day of every month. The resulting discrete sample p-p plot, its continuous counterpart, and the concomitant HWM index are all depicted in Figure 2.2.

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 2.1 Maximum water levels of the river Meuse at Borgharen Dorp, in millimetres, measured on every last day of every month.
Source: Koninklijk Nederlands Meteorologisch Instituut, personal correspondence (see Hinloopen, 1997, Chapter 5).

Figure 2.2 Discrete p-p plot for water levels of the river Meuse at Borgharen Dorp (The Netherlands), 1990 versus 1993 (Panel a), the corresponding continuous p-p plot (Panel b), the concomitant HWM index (Panel c), and the resulting d-d plot (Panel d).



2.2. Computation

Note that the value of the shaded area in Panel c of Figure 2.2, which equals 7/12, is straightforward to obtain. For samples with ties this is not necessarily the case. Within-sample ties affect the number of coordinates that make up the p-p plot. As such they have no effect on the possible slope of the p-p plot. Between-sample ties may induce the p-p plot to have linear pieces with any positive slope. This complicates the exact computation of the HWM index. For computing the HWM index we introduce the diagonal-deviation plot: the projection of the p-p plot onto the diagonal.[3] This means that each point ($F_1(z)$, $F_2(z)$) on the p-p plot is projected on the average probability $p(z) \equiv (F_1(z) + F_2(z))/2$. Because the length of the projection vector equals $d(z) \equiv |F_1(z) - F_2(z)|/\sqrt{2}$, we arrive at the following definition:

Definition 2.2. The diagonal-deviation plot is the projection of the p-p plot onto the diagonal:

$$z \mapsto \begin{bmatrix} p(z) \\ d(z) \end{bmatrix},\ p(z) \equiv \left(F_1(z) + F_2(z)\right)/2,\ d(z) \equiv |F_1(z) - F_2(z)|/\sqrt{2}.$$

---

[3] Use of this projection is not necessary for computing the exact value of the 2-sample HWM index but it allows for a straightforward extension of the HWM index to the simultaneous comparison of $k > 2$ samples (see Section 3).

The d-d plot is obtained by first projecting the coordinates of the discrete p-p plot onto the diagonal. This gives the coordinates of the d-d plot: $D \equiv (p(z_i), d(z_i))$, $i = 1, 2, ..., m$, where $m \leq n_1 + n_2$, and $(p(0), d(0)) \equiv (0,0)$. Next, linearly interpolate between these coordinates to obtain the d-d plot. The value of the HWM index then corresponds to the surface between the d-d plot line and the horizontal zero axis multiplied by the projection scaling factor $PS(2) = \sqrt{2}$ (see Panel d in Figure 2.2).

Computation of the HWM index requires the exact location of all intersections of the underlying p-p plot with the diagonal. Diagonal cutting points are all points $(F_1(z), F_2(z))$ for which $d(z) = 0$. Their location follows from the following lemma (the proof of which is in Appendix A, Section 6.2):

Lemma 2.1. When the p-p plot cuts the diagonal between points z_i and z_{i+1}, $z_i \in \mathbf{z}$, then

$$p_h = F_{1,n_1}(z_i) + \frac{(F_{1,n_1}(z_{i+1}) - F_{1,n_1}(z_i))(F_{1,n_1}(z_i) - F_{2,n_2}(z_i))}{(F_{2,n_2}(z_{i+1}) - F_{2,n_2}(z_i)) - (F_{1,n_1}(z_{i+1}) - F_{1,n_1}(z_i))}$$

if at probability p_h we have that $F_{1,n_1}(z_i) \geq F_{2,n_2}(z_i)$ and $F_{1,n_1}(z_{i+1}) \leq F_{2,n_2}(z_{i+1})$, that is, the diagonal is "cut from below", and

$$p_h = F_{2,n_2}(z_i) + \frac{(F_{2,n_2}(z_{i+1}) - F_{2,n_2}(z_i))(F_{2,n_2}(z_i) - F_{1,n_1}(z_i))}{(F_{1,n_1}(z_{i+1}) - F_{1,n_1}(z_i)) - (F_{2,n_2}(z_{i+1}) - F_{2,n_2}(z_i))}$$

if at probability p_h we have that $F_{1,n_1}(z_i) \leq F_{2,n_2}(z_i)$ and $F_{1,n_1}(z_{i+1}) \geq F_{2,n_2}(z_{i+1})$, that is, the diagonal is "cut from above".

Obviously, the p-p plot can cut the diagonal at points that are not in D. Let L be the set with these diagonal cutting points consisting of l entries and let $I = D \cup L \equiv (p_i^*, d_i^*)$, $i = 1, 2, ..., m + l$, be ordered on p_i^*. The HWM index can then be calculated exactly under all circumstances (the proofs of all propositions are in Appendix A, Section 6.3):

Proposition 2.1.

$$HWM(F_{1,n_1}, F_{2,n_2}) = PS(2)S(n_1, n_2) \times \sum_{i=1}^{m+l} \left[\frac{\left(p_i^* - p_{i-1}^*\right)}{2} \left(2 \times \max\{d_i^*, d_{i-1}^*\} - \left|d_i^* - d_{i-1}^*\right|\right) \right].$$

2.3. Hypothesis testing

Being able to compute the value of the HWM index exactly allows for the simulation of significance tables under varying circumstances. For a special case however the exact finite sample distribution of the HWM index is known.

2.3.1. Finite sample distribution

Consider the set of continuous distribution functions Ξ_2. Observe that $\Xi_1 \subset \Xi_1$. Also, because functions belonging to Ξ_2 are strictly increasing on their support, there are no mass points. That is, there are no ties. Indeed, the HWM index coincides with the L_1–version of the Fisz-Cramér-von Mises statistic (Schmidt and Trede, 1995) for distributions belonging to Ξ_2 and for two samples of equal size n. We thus introduce:

Assumption 2.A1. $F_1, F_2 \in \Xi_2$,

Assumption 2.A2. $n_1 = n_2 = n.$

Proposition 2.2. Under A1 – A2, the HWM index coincides with the L_1–FCvM statistic:

$$\frac{(n_1 n_2)^{1/2}}{(n_1 + n_2)^{3/2}} \sum_{i=1}^{n_1+n_2} \left| F_{1,n_1}(z_i) - F_{2,n_2}(z_i) \right|.$$

Hinloopen and Wagenvoort (2010) derive the analytical solution for the probability density function of the L_1–FCvM statistic under H_0 and A1 – A2. For these cases, the exact critical percentiles of the HWM index are thus known. Indeed, the diagonal entries of Tables B.5 through B.8 in Appendix B correspond to these percentiles for two balanced samples up to $n = 20$.

High values of the HWM index imply a low probability that the underlying samples are drawn from the same distribution. For instance, for the data in Table 2.1 the HWM index value is 0.2381, which is substantially smaller than the critical percentiles at common significance levels. Although the southern part of The Netherlands was flooded in 1993, considering the entire year shows that the monthly maximum water levels of the river Meuse at Borgharen Dorp in 1993 were not significantly different from the levels in 1990.

2.3.2. Limiting distribution

For balanced samples without ties Schmid and Trede (1995) note that the limiting distribution of the L_1–FCvM statistic corresponds to the limiting distribution of the L_1–norm of a Brownian bridge. Proposition 2.2 implies that the same holds for the HWM index. The analytical expression for the L_1–norm of a Brownian bridge is derived by Johnson and Killeen (1983). They also tabulate the concomitant critical values. It appears that the critical values of the HWM index converge rapidly to their limiting values (see Table 1 in Hinloopen and Wagenvoort, 2010). Indeed, although for finite samples the critical percentiles of the limiting distribution differ from the exact values, this will not lead to differences in the rejection of the underlying hypothesis (see Hinloopen and Wagenvoort, 2010).

2.3.3. Ties

Ties can be present in any sample, even for continuous population distributions due to rounding. Within-sample ties reduce the number of coordinates that constitute the p-p plot. Between-sample ties allow the p-p plot to remain closer to the diagonal. Ties therefore affect every EDF that quantifies a p-p plot. One way of dealing with ties is to use a randomized tie-breaking procedure (Dufour (1995), Dufour and Kiviet, 1998). Let U_i, $i = 1, ..., n_1 + n_2$ be a random sample of $n_1 + n_2$ observations from a uniform continuous distribution. The observations $Z = X_1 \cup X_2$ can then be arranged following the order:

$$(Z_i, U_i) < (Z_j, U_j) \Leftrightarrow Z_i < Z_j \text{ or } (Z_i = Z_j \text{ and } U_i < U_j), \tag{2.3}$$

which results in $n_1 + n_2$ different order statistics. The test statistic is then computed for these $n_1 + n_2$ different order statistics rather than the $q < n_1 + n_2$ order statistics from the original samples.

Alternatively the q order statistics are used and the critical values in Appendix B are applied at the possible cost of a small size distortion in the critical area. Indeed, within-sample ties possibly increase the value of the HWM index whereas between-sample ties possibly reduce it. To assess the impact of ties on the distribution of the HWM index we have to rely on numerical simulations because assumption 2.A1 is violated.

percentile	90	95	97.5	99
domain				
{0, 1}	-0.047	-0.027	-0.018	-0.006
{0, 1, 2}	-0.040	-0.022	-0.014	-0.006
{0, 1, 2, 3}	-0.023	-0.013	-0.007	-0.002
{0,…, 4}	-0.023	-0.012	-0.009	-0.003
{0,…, 5}	-0.018	-0.011	-0.005	-0.003
{0,…, 6}	-0.016	-0.009	-0.002	0.001
{0,…, 7}	-0.012	-0.007	-0.004	-0.002
{0,…, 8}	-0.011	-0.006	-0.003	-0.001
{0,…, 9}	-0.012	-0.007	-0.004	-0.001
{0,…, 10}	-0.008	-0.005	-0.003	-0.002
{0,…, 11}	-0.010	-0.008	-0.004	-0.001
{0,…, 12}	-0.006	-0.004	-0.004	-0.001
{0,…, 13}	-0.004	-0.006	-0.003	-0.002
{0,…, 14}	-0.005	-0.005	-0.003	-0.001
{0,…, 15}	-0.008	-0.006	-0.006	-0.002

Table 3.2 Size distortions due to ties in balanced samples.
Notes: All samples consist of 50 entries which are drawn from the integer domains in the first column, where all entries have equal probability. The resulting simulated distribution for each row consists of 10,000 independent HWM index values.

As the influence of ties on the distribution of the HWM index turns out to be negligible in most cases, we report only the simulation results for an extreme situation in that all observations constitute a tie. Table 3.2 lists the size distortions of the exact critical percentiles derived by Hinloopen and Wagenvoort (2010). This distortion is defined as prob[$HWM > cv$] – cp, where cv is the critical value of the HWM index under A1 – A2 and cp is the concomitant probability (i.e. 10%, 5%, 2.5% and 1%). Clearly, what matters is the number of classes that underlie the sample EDFs. For instance, if there are only 2 classes, 0 and 1 say, the 90th percentile is 0.400 while without ties it equals 0.500, yielding a size distortion of -0.047. But size distortions fall rapidly when the number of classes increases. With as little as 7 classes the simulated 90th percentile is already 0.485 and the size distortion is one percentage point only. It thus seems that for the HWM index the effect of within-sample ties and between-sample ties cancel out when the number of classes is sufficiently high. Because many applications will involve less than 100% ties (the only other study we know of that considers the effect of ties, Scholz and Stephens (1987), reports simulations up to situations where 60% of all observations constitute a tie), we conjecture that the critical values shown in Appendix B are accurate under most circumstances where assumption 2.A1 is violated.

Perhaps more importantly, the HWM index is less sensitive to ties than the L_1–FCvM statistic. This is because the HWM index is invariant to the position of a tie whereas the FCvM statistic is not. For instance, let X_1={1, 2, 3}, and X_2= {1.5, 1.5, 4} with respective discrete EDFs $F_{1,3}$ and $F_{2,3}$. If the entries in X_2 are rounded upwards, yielding $\bar{X}_2 = \{2,\ 2,\ 4\}$ and $F_{2,\bar{3}}$, a between-sample tie arises at the second entry of both samples. This gives HWM$(F_{1,3}, F_{2,\bar{3}})$ = 0.2041 and $L_1 - \text{FCvM}(F_{1,3}, F_{2,\bar{3}}) = 0.1361$. Alternatively the entries are rounded downwards, giving us $\underline{X}_2 = \{1, 1, 4\}$ and $F_{2,\underline{3}}$. A between-sample tie now occurs at the first entry of both samples. This leaves the HWM index unaffected, HWM$(F_{1,3}, F_{2,\underline{3}})$ = 0.2041 while it does influence the FCvM statistic: $L_1 - \text{FCvM}(F_{1,3}, F_{2,\underline{3}}) = 0.2722$. That is, the HWM index takes the same value, whether the between-sample tie occurs at the first or second entry. The HWM index is unaffected because it is based on a continuous p-p plot where mass points are removed by spreading probability uniformly between two order statistics. In contrast, the FCvM statistic changes value depending on the position of the tie as the FCvM statistic is based on the discrete p-p plot. This makes the FCvM statistic a less robust statistic since the augmented samples $\overline{X}_2$ and $\underline{X}_2$ are 'equally close' to sample X_1.

2.3.4. Unbalanced samples

To assess the influence of unbalanced samples we again have to revert to numerical simulations. Tables B.5 through B.8 in Appendix B (Section 7.1) list the simulated percentiles of the HWM index under H_0 for all possible unbalanced samples up to $n = 20$. Clearly, sample size correction factor (2) adequately deals with unbalanced samples. For larger samples the simulated percentiles suggest a simple rule of thumb: in case of unbalanced samples that are not in Appendix B use the analytical value of the HWM index for the largest sample size. Even in an extreme case where one sample consists of 3 entries while the other has 20 entries this yields a small approximation error.

3. THE K-SAMPLE HWM INDEX

For extending the HWM index to the simultaneous comparison of $k > 2$ samples we first have to introduce the k-sample p-p plot:

Definition 3.3. The k-sample p-p plot depicts for every domain value z from the joint support of $F_1, ..., F_k$ the percentiles of one distribution relative to the others:

$$z \mapsto \begin{bmatrix} F_1(z) \\ ... \\ F_k(z) \end{bmatrix}.$$

Obviously k-sample p-p plots are mappings from [0, 1] to $[0, 1]^{k\text{-}1}$. Hypothesis H_0 then extends to: H_0^*: $F_1 = \ldots = F_k$. And the generalized assumptions 2.A1 and 2.A2 respectively read as:

Assumption 3.A1*. $F_1, ..., F_k \in \Xi_2$,

Assumption 3.A2*. $n_1 = ... = n_k = n$.

3.1. Definition

The k-dimensional HWM index is defined as the surface between the k-dimensional p-p plot and the 'diagonal' that cuts all 2-dimensional spaces in equal halves. The coordinates of this diagonal are given by the *k*-dimensional vector $(\tilde{p}(x),\ldots,\tilde{p}(x))$, where $\tilde{p}(x)$ is the average probability at *x*. For *k* > 2 this surface is not uniquely defined however. It depends on the point of the diagonal on which probabilities are projected. An obvious candidate is to take the shortest distance between the p-p plot and the diagonal, which implies that each point of the p-p plot is projected on the concomitant average probability.[4] For characterizing this distance we use the Mahalanobis distance:

Definition 3.4.

$$HWM(F_1,F_2,\ldots,F_k) \equiv PS(k)S(n_1,\ldots,n_k)\times\int_{-\infty}^{\infty}\sqrt{\sum_{j=1}^{k}\left(\tilde{F}_j(x)-\tilde{p}(x)\right)^2}\,d\tilde{p}(x),$$

where $\tilde{F}_j(x)$ is the continuous analogue of the possibly discrete EDF of X_j, $\tilde{p}(x)=\frac{1}{k}\sum_{j=1}^{k}\tilde{F}_j(x)$ is the average probability at *x*, where $S(n_1,\ldots,n_k)$ is a multi-sample scaling factor, and where *PS*(*k*) is a factor that scales the projection. In particular we generalize (2) to:

$$S(n_1,\ldots,n_k)=\frac{\left(\Pi_{j=1}^{k}n_j\right)^{\frac{1}{k}}}{\left(\sum_{j=1}^{k}n_j\right)^{\frac{1}{2}}}. \tag{3.4}$$

The projection scaling factor *PS*(*k*) maintains the correspondence between the surface below the d-d plot and the surface between the k-dimensional p-p plot and the diagonal. It equals the length of the diagonal of the k-dimensional p-p plot:

$$PS(k)=\sqrt{k}. \tag{3.5}$$

Note that the 2-sample version of the HWM index in Definition 2.1 corresponds to Definition 3.4 with *k* = 2.

The properties of the 2-sample HWM index carry over to the k-sample version:

Property 3.P1* *(equality)*: $HWM(F_1,\ldots,F_k)=0 \Leftrightarrow \forall z\in[a,b]: F_1(z)=\ldots=F_k(z)$.

Property 3.P2* *(order irrelevance)*: $HWM(F_1,\ldots,F_k)=HWM(G(F_1,\ldots,F_k))$, where *G(·)* is any perturbation of the order of its entries.

3.2. Computation

Computing the *k*-sample HWM index involves again the construction of the underlying d-d plot. In case the p-p plot is projected on the average probabilities, the coordinates of the corresponding d-d plot for any $z_i \in \mathbf{z}$, are:

$$p(z_i)=\frac{1}{k}\sum_{j=1}^{k}F_j(z_i), \tag{3.6}$$

$$d(z_i)=\sqrt{\sum_{j=1}^{k}\left(F_j(z_i)-p(z_i)\right)^2}, \tag{3.7}$$

---

[4] We also considered alternatives, such as the minimum probability. This led to much more variability in the HWM index which, we think, makes it less suitable for hypotheses testing.

$i = 1, \ldots, m$, $m \le \sum_{j=1}^{k} n_j$, and $(p(0), d(0)) \equiv (0,0)$. Obviously, Proposition 2.1 applies such that the k-sample HWM index can be computed as follows:

$$HWM(F_{1,n_1}, \ldots, F_{k,n_k}) = PS(k)S(n_1, \ldots, n_k) \times \sum_{i=1}^{m+l} \left[\frac{\left(p_i^* - p_{i-1}^*\right)}{2} \left(2 \max\{ d_i^*, d_{i-1}^* \} - \left| d_i^* - d_{i-1}^* \right| \right) \right],$$

in the notation of Proposition 2.1.

3.3. Hypotheses testing

An exact formulation is not known for the finite sample distribution of any of the existing k-sample EDF tests with $k > 2$ (Kiefer, 1959). However, as the HWM index is distribution free under assumptions 3.A1* – 3.A2*, we can simulate its distribution. The concomitant critical percentiles are in Tables B.9 through B.12 in Appendix B for $k = 3, \ldots, 15$ for sample sizes up to the point of convergence.

4. LONG-TERM SOVEREIGN DEBT CREDIT RATINGS

The performance of the HWM test should be measured against its ability to discriminate between samples that are not drawn from the same population distribution. As none of the existing EDF tests dominates all other tests under all circumstances, in principle all tests must be considered at all times. However, in some cases a particular test is known to have the best power. For instance, when samples are drawn from distributions with different tails, the Anderson-Darling (AD) test outperforms all alternatives as it puts more weight on extremities. The HWM test on the other hand is expected to outperform all other tests when the p-p plot remains relatively close to the diagonal because it is the only EDF test that assigns equal weight to all distances between the respective distributions.

Rating agency	Fitch	Moody's	S&P
Credit ratings			
AAA, AA+	25	23	23
AA, AA-	6	13	13
A+, A	11	9	8
A-, BBB+	5	10	8
BBB, BBB-	26	16	21
BB+, BB	10	13	13
BB-, B	9	13	11
B-, CCC+	6	4	4
CCC, CCC-	1	0	0
CC+, CC	0	1	1
	Value test statistic		
HWM	0.163		
AD	-1.690		
Number of observations	$n_1 = n_2 = n_3 = 80$		

Table 4.3 Share (in %) of countries ordered according to their long-term sovereign debt credit ratings, September 7, 2011.
Notes: The upper 90th critical percentile is 0.701 and 1.309 for HWM and AD respectively.

As an illustration consider the long-term sovereign debt credit ratings of the main rating agencies Fitch, Moody's, and Standard & Poor's (S&P). On 7 Sept. 2011, in total 80 countries were rated by each agency (IMF, 2011). Out of 22 possible ratings, each of the 20 ratings from AAA to CC was given to at least one country. Table 4.3 contains the rating distributions, where each ‘bin’ consists of two ratings. For example, Fitch rated 25% of the 80 countries AAA or AA+. Obviously, there are many within and between sample ties across the rating distributions. Both the HWM and the Anderson-Darling test do not reject the null-hypothesis that the ratings are drawn from the same underlying distribution at the 10 percent or higher confidence level; there are no differences in outcome of the rating methodologies across the three rating agencies.

In order to compare the power of HWM and AD, we first use the rating distributions as the basis for a Monte Carlo simulation. For that we randomly draw sub-samples of 10 or 50 observations from the Moody’s and S&P rating data. A third sub-sample is drawn from the uniform distribution with integer domain [1, 2, …, 20] rather than from the Fitch rating distribution. Next we compute both HWM and AD and determine the fraction of rejections in 10,000 trials. The results are in Table 4.4.

Number of observations $n_1 = n_2 = n_3$	Test	
	HWM	AD
	Share of rejections at the 10% confidence level	
10	28	39
50	83	95
	Share of rejections at the 5% confidence level	
10	17	28
50	76	92
	Share of rejections at the 2.5% confidence level	
10	11	19
50	66	87
	Share of rejections at the 1% confidence level	
10	5	11
50	54	78

Table 4.4 Power assessment in a Monte Carlo experiment.
Notes: Observations are randomly drawn from the Moody’s and S&P rating distributions. A third sub- sample is drawn from the uniform distribution on the integer domain [1, 2,…,20]. Results are based on 10,000 trials.

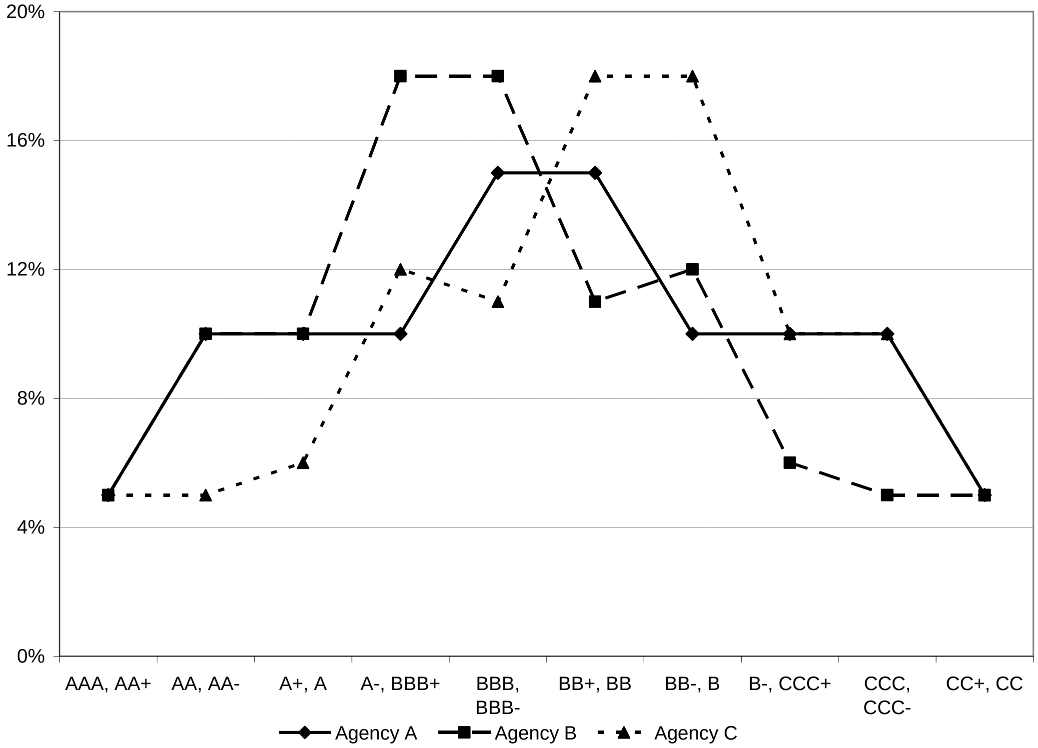
AD has more power than HWM to discriminate between the three sub-samples. In small samples of 10 observations, AD rejects H_0 in nearly twice as many cases as HWM. Differences in power are smaller for larger samples. For example, at the 10% confidence level HWM rejects H_0 in 83% of the cases with sub-samples of 50 observations compared to a rejection rate of 95% for AD.

For the simulation underlying Table 4.4, observations are drawn independently from each distribution. However, if observations are not drawn independently, HWM can have more power than AD. For instance, if we fix the relative frequencies of the 10 bins in Table 4.3 such that distributions differ significantly, but relatively little at the extremities, examples can be found where HWM outperforms AD. Figure 4.3 depicts such a situation. It depicts fictitious rating distributions for three fictitious rating agencies.[5] If these distributions are used when drawing samples of 100 observations, a different picture arises: at the 5% confidence level, HWM rejects H_0 in 42.9% of all cases, whereas AD rejects H_0 in 38.6% only.

---

[5] In particular, the probabilities over the respective 10 bins for the three fictitious rating agencies are (in percentages): {5, 10, 10, 10, 15, 15, 10, 10, 10, 5}, {5, 10, 10, 18, 18, 11, 12, 6, 5, 5} and {5, 5, 6, 12, 11, 18, 18, 10, 10, 5}.

Figure 4.3 Fictitious probabilities with which samples are drawn from the 10 rating bins for three fictitious rating agencies, yielding HMW to have higher power than AD at the 5% confidence level.



5. Discussion and conclusions

The quantification of the p-p plot we propose assigns equal weight to all distances between the respective distributions. For two balanced samples without ties it is shown that the HWM index coincides with the L_1–FCvM statistic (see Schmidt and Trede (1995); see also Hinloopen and Wagenvoort, 2010). The HWM index differs from the L_1–FCvM statistic when there are ties in that it is invariant to the position of the tie in the sequence of order statistics. This makes it a more robust statistic.

We further show how the HWM index can be extended to a homogeneity test involving $k > 2$ samples. For that we have introduced the k-sample pp-plot. We also show how the k-sample HWM index can be computed exactly under all circumstances. In so doing we have introduced the d-d plot: the projection of the p-p plot onto the diagonal.

Both ties and unbalanced samples appear to affect the distribution of the 2-sample HWM index only mildly. However, ties can have a non-negligible effect on the distribution of the k-sample HWM index for $k > 2$, even if the underlying population distributions consist of many classes. This is because the probability of a between-sample tie increases with k relative to the probability of a within-sample tie, and different ties have an opposite effect on the value of the HWM index.

In sum, as none of the existing EDF tests, including the HWM test, outperforms all other tests under all circumstances, they should all be considered simultaneously so as to avoid type-II errors as much as possible.

APPENDIX

Appendix A Proofs

A.1. Proofs of properties

Proof of Property P1*. Let [a,b] be the common support of $\tilde{F}_j(z)$, j = 1, …, k. $HWM(F_1, F_2, ..., F_k) = 0 \Leftrightarrow$ because of the continuity of both $\tilde{F}_j(z)$ and $\tilde{p}(z)$: $\sqrt{\sum_{j=1}^{k}\left(\tilde{F}_j(z) - \tilde{p}(z)\right)^2} = 0 \;\forall\; z \in [a,b] \;\Leftrightarrow\; \tilde{F}_1(z) = \tilde{p}, ..., \tilde{F}_k(z) = \tilde{p},\; \forall\; z \in [a,b]$ $\Leftrightarrow F_1(z) = ... = F_k(z), \forall\; z \in [a,b]$.

QED

Proof of Property P2*. $HWM(F_1, F_2, ..., F_k) \equiv PS(k)S(n_1, ..., n_k) \times$ $\int_{-\infty}^{\infty} \sqrt{\sum_{j=1}^{k}\left(\tilde{F}_j(z) - \tilde{p}(z)\right)^2}\, d\tilde{p}(z) \;=\; PS(k)S(n_1, ..., n_k) \int_{-\infty}^{\infty} \sqrt{G\left(\sum_{j=1}^{k}\left(\tilde{F}_j(z) - \tilde{p}(z)\right)^2\right)} d\tilde{p}(z)$ $= HWM(G(F_1, ..., F_k))$, where $G(\cdot)$ is any perturbation of the order of its entries.

QED

A.2. Proofs of Lemma 2.1

A p-p plot that cuts the diagonal from below at probability p_h between observations z_i and z_{i+1} (i.e. $F_{1,n_1}(z_i) \geq F_{2,n_2}(z_i)$ and $F_{1,n_1}(z_{i+1}) \leq F_{2,n_2}(z_{i+1})$) is illustrated in Figure A.4. The diagonal runs from $(F_{1,n_1}(z_i), F_{1,n_1}(z_i))$ to $(F_{1,n_1}(z_{i+1}),\; F_{1,n_1}(z_{i+1}))$ while the p-p plot connects $(F_{1,n_1}(z_i), F_{2,n_1}(z_i))$ with $(F_{1,n_1}(z_{i+1}), F_{2,n_2}(\, z_{i+1}))$. Let $a = p_h - F_{1,n_1}(z_i)$, $b = F_{1,n_1}(z_i) - F_{2,n_2}(z_i)$, $c = F_{1,n_1}(z_{i+1}) - F_{1,n_1}(z_i)$, and d = $F_{2,n_2}(z_{i+1}) - F_{2,n_2}(z_i)$. From $TAN(t)$ = d/c = (a + b)/a we obtain that a = bc/(d − c). Hence, p_h follows. A similar argument applies in case the diagonal is cut from above.

QED

A.3. Proofs of propositions

Proof of Proposition 2.1. The surface between the p-p plot and the diagonal equals the surface below the d-d plot multiplied by $\sqrt{k}$. The surface below the d-d plot between (p^*_{i-1}, d^*_{i-1}) and (p^*_i, d^*_i) equals $(p^*_i - p^*_{i-1})d^*_{i-1} + (p^*_i - p^*_{i-1})(d^*_i - d^*_{i-1})/2$ if $d^*_i \geq d^*_{i-1}$, and $(p^*_i - p^*_{i-1})d^*_i + (p^*_i - p^*_{i-1})(d^*_{i-1} - d^*_i)/2$ if $d^*_i \leq d^*_{i-1}$. The proposition then follows.

QED

Proof of Proposition 2.2. Under A1 – A2, the contribution of any two points z_i and z_{i+1} to the L_1-FCvM statistic equals

$$\frac{\left|F_2(z_i) - F_1(z_i)\right| + \left|F_2(z_{i+1}) - F_1(z_{i+1})\right|}{2\sqrt{2n}}.$$

The surface between the diagonal and the p-p plot from point z_i to z_{i+1} multiplied by the scaling factor $S(n_1, n_2) = n/\sqrt{2n}$ equals

$$\left[\frac{n}{\sqrt{2n}}\right]\times\left[\frac{F_2(z_{i+1})-F_2(z_i)}{2n}+\frac{F_2(z_i)-F_1(z_{i+1})}{n}+\frac{F_1(z_{i+1})-F_1(z_i)}{2n}\right]=$$
$$\frac{F_2(z_i)-F_1(z_i)+F_2(z_{i+1})-F_1(z_{i+1})}{2\sqrt{2n}},$$

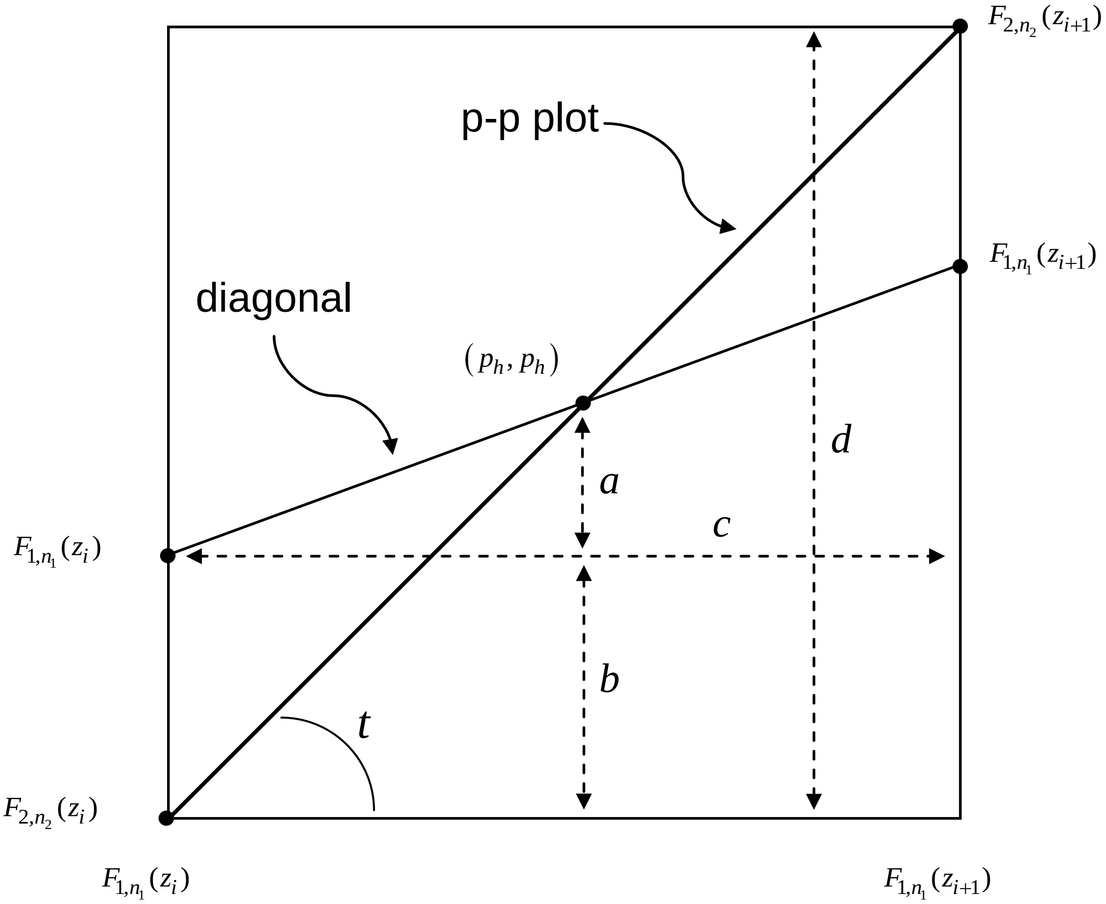
when $F_2(z_i)\geq F_1(z_i)$ and $F_2(z_{i+1})\geq F_1(z_{i+1})$, i.e. the p-p plot is above the diagonal, and equals

$$\left[\frac{n}{\sqrt{2n}}\right]\times\left[\frac{F_2(z_{i+1})-F_2(z_i)}{2n}+\frac{F_1(z_i)-F_2(z_{i+1})}{n}+\frac{F_1(z_{i+1})-F_1(z_i)}{2n}\right]=$$
$$\frac{F_1(z_i)-F_2(z_i)+F_1(z_{i+1})-F_2(z_{i+1})}{2\sqrt{2n}},$$

when $F_2(z_i)\leq F_1(z_i)$ and $F_2(z_{i+1})\leq F_1(z_{i+1})$, i.e. the p-p plot is below the diagonal. The contribution of any two points z_i and z_{i+1} to the value of the HWM index is thus identical to their contribution to the value of the L_1-FCvM statistic.

QED

Figure A.4 A p-p plot cutting the diagonal “from below” at point (p_h, p_h).



Appendix B Critical percentiles

B.1. *k* = 2, unbalanced samples, no ties

n	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	0.612																	
4	0.546	0.530																
5	0.502	0.522	0.538															
6	0.511	0.516	0.495	0.529														
7	0.518	0.513	0.512	0.514	0.515													
8	0.492	0.510	0.510	0.501	0.518	0.500												
9	0.528	0.508	0.498	0.498	0.501	0.505	0.511											
10	0.516	0.507	0.511	0.506	0.505	0.502	0.508	0.514										
11	0.501	0.506	0.504	0.508	0.512	0.509	0.506	0.502	0.494									
12	0.516	0.505	0.501	0.500	0.508	0.506	0.504	0.506	0.502	0.510								
13	0.503	0.505	0.512	0.498	0.504	0.502	0.503	0.512	0.504	0.500	0.505							
14	0.495	0.504	0.498	0.505	0.496	0.504	0.508	0.513	0.504	0.508	0.509	0.499						
15	0.509	0.503	0.512	0.506	0.502	0.505	0.502	0.504	0.498	0.502	0.508	0.510	0.505					
16	0.497	0.503	0.508	0.500	0.504	0.505	0.500	0.500	0.505	0.506	0.506	0.501	0.500	0.508				
17	0.507	0.501	0.501	0.510	0.508	0.508	0.508	0.502	0.512	0.502	0.502	0.501	0.504	0.507	0.499			
18	0.510	0.503	0.506	0.511	0.506	0.498	0.507	0.507	0.505	0.505	0.498	0.501	0.508	0.495	0.502	0.500		
19	0.499	0.503	0.508	0.506	0.500	0.502	0.499	0.499	0.506	0.499	0.506	0.506	0.503	0.506	0.504	0.502	0.499	
20	0.501	0.502	0.505	0.507	0.510	0.509	0.501	0.503	0.513	0.505	0.502	0.502	0.505	0.511	0.500	0.495	0.509	0.506

Table B.5 Simulated critical values at percentile 90 for unbalanced samples without ties.

Notes: The underlying distribution consists of 10,000 independent HWM index values that are computed for samples that are drawn from a standard normal distribution. The joint number of sample entries must be at least 6 for percentile 90 to be defined properly.

n	3	4	5	6	7	8	9	10	11
4	0.655	0.619							
5	0.593	0.596	0.601						
6	0.550	0.581	0.606	0.577					
7	0.587	0.570	0.581	0.599	0.592				
8	0.577	0.587	0.596	0.582	0.587	0.594			
9	0.583	0.601	0.578	0.586	0.583	0.575	0.589		
10	0.608	0.592	0.584	0.581	0.583	0.580	0.595	0.581	
11	0.582	0.584	0.590	0.584	0.593	0.587	0.578	0.583	0.591
12	0.570	0.577	0.595	0.583	0.584	0.593	0.588	0.585	0.590
13	0.580	0.582	0.599	0.572	0.578	0.578	0.581	0.588	0.586
14	0.569	0.576	0.576	0.580	0.584	0.584	0.594	0.587	0.596
15	0.580	0.565	0.590	0.587	0.582	0.590	0.589	0.588	0.578
16	0.569	0.587	0.586	0.588	0.591	0.586	0.583	0.587	0.580
17	0.579	0.582	0.573	0.585	0.589	0.593	0.587	0.587	0.592
18	0.584	0.578	0.593	0.589	0.588	0.583	0.590	0.585	0.587
19	0.583	0.579	0.577	0.581	0.576	0.590	0.579	0.579	0.582
20	0.569	0.589	0.585	0.579	0.591	0.589	0.584	0.587	0.593

n	12	13	14	15	16	17	18	19	20
12	0.578								
13	0.591	0.581							
14	0.590	0.586	0.594						
15	0.591	0.594	0.589	0.590					
16	0.586	0.592	0.586	0.591	0.586				
17	0.585	0.575	0.582	0.592	0.585	0.580			
18	0.584	0.577	0.581	0.583	0.576	0.580	0.593		
19	0.583	0.587	0.587	0.585	0.591	0.584	0.580	0.585	
20	0.586	0.583	0.583	0.586	0.589	0.583	0.581	0.598	0.585

Table B.6 Simulated critical values at percentile 95 for unbalanced samples without ties.

Notes: The underlying distribution consists of 10,000 independent HWM index values that are computed for samples that are drawn from a standard normal distribution. The joint number of sample entries must be at least 7 for percentile 95 to be defined properly.

n	3	4	5	6	7	8	9	10	11
5	0.685	0.671	0.664						
6	0.629	0.645	0.661	0.674					
7	0.656	0.627	0.659	0.642	0.668				
8	0.615	0.663	0.658	0.656	0.656	0.656			
9	0.639	0.647	0.657	0.650	0.646	0.657	0.642		
10	0.658	0.651	0.657	0.645	0.667	0.659	0.653	0.648	
11	0.641	0.662	0.657	0.657	0.673	0.660	0.663	0.645	0.649
12	0.645	0.650	0.658	0.653	0.651	0.662	0.658	0.662	0.655
13	0.661	0.673	0.658	0.635	0.648	0.655	0.660	0.660	0.655
14	0.636	0.661	0.658	0.645	0.661	0.665	0.662	0.670	0.662
15	0.650	0.630	0.658	0.646	0.655	0.662	0.659	0.653	0.660
16	0.635	0.643	0.647	0.653	0.650	0.659	0.659	0.660	0.640
17	0.642	0.635	0.643	0.661	0.646	0.656	0.660	0.655	0.670
18	0.653	0.653	0.659	0.648	0.659	0.654	0.650	0.662	0.673
19	0.635	0.646	0.644	0.643	0.655	0.664	0.658	0.649	0.651
20	0.646	0.662	0.660	0.657	0.667	0.663	0.651	0.658	0.666
n	12	13	14	15	16	17	18	19	20
12	0.646								
13	0.657	0.671							
14	0.664	0.656	0.675						
15	0.666	0.670	0.663	0.663					
16	0.668	0.662	0.659	0.661	0.663				
17	0.658	0.654	0.654	0.661	0.664	0.661			
18	0.671	0.657	0.657	0.657	0.647	0.655	0.667		
19	0.662	0.669	0.651	0.655	0.666	0.654	0.658	0.662	
20	0.658	0.659	0.656	0.664	0.662	0.651	0.658	0.672	0.656

Table B.7 Simulated critical values at percentile 97.5 for unbalanced samples without ties.

Notes: The underlying distribution consists of 10,000 independent HWM index values that are computed for samples that are drawn from a standard normal distribution. The joint number of sample entries must be at least 8 for percentile 97.5 to be defined properly.

n	3	4	5	6	7	8	9	10	11
5		0.745	0.727						
6	0.707	0.71	0.716	0.722					
7	0.725	0.684	0.708	0.728	0.745				
8	0.739	0.714	0.702	0.733	0.729	0.719			
9	0.694	0.74	0.737	0.738	0.74	0.743	0.746		
10	0.709	0.718	0.73	0.742	0.745	0.738	0.747	0.76	
11	0.675	0.74	0.725	0.746	0.766	0.734	0.73	0.728	0.727
12	0.689	0.722	0.72	0.722	0.726	0.73	0.756	0.743	0.744
13	0.701	0.74	0.745	0.727	0.738	0.749	0.753	0.75	0.743
14	0.711	0.724	0.74	0.732	0.727	0.745	0.755	0.759	0.754
15	0.692	0.711	0.736	0.736	0.739	0.742	0.747	0.746	0.74
16	0.702	0.727	0.715	0.718	0.749	0.74	0.717	0.744	0.74
17	0.705	0.714	0.728	0.743	0.72	0.737	0.737	0.738	0.767
18	0.713	0.729	0.725	0.727	0.737	0.735	0.726	0.758	0.76
19	0.696	0.718	0.723	0.731	0.746	0.749	0.73	0.727	0.739
20	0.7	0.73	0.74	0.734	0.755	0.747	0.734	0.762	0.751

n	12	13	14	15	16	17	18	19	20
12	0.731								
13	0.737	0.762							
14	0.741	0.743	0.756						
15	0.757	0.773	0.743	0.736					
16	0.75	0.753	0.734	0.753	0.74				
17	0.741	0.743	0.736	0.747	0.76	0.741			
18	0.758	0.751	0.746	0.75	0.738	0.744	0.75		
19	0.737	0.759	0.736	0.747	0.756	0.747	0.747	0.764	
20	0.745	0.734	0.748	0.742	0.743	0.739	0.737	0.756	0.743

Table B.8 Simulated critical values at percentile 99 for unbalanced samples without ties.

Notes: The underlying distribution consists of 10,000 independent HWM index values that are computed for samples that are drawn from a standard normal distribution. The joint number of sample entries must be at least 9 for percentile 99 to be defined properly.

B.2. *k* = 3, ..., 15, balanced samples, no ties

n	3	4	5	6	7	8	9	10	50	100
K										
3	0.720	0.706	0.706	0.704	0.710	0.696	0.703	0.705	0.692	0.701
4	0.841	0.840	0.843	0.843	0.840	0.841	0.838	0.837	0.839	0.841
5	0.948	0.951	0.951	0.949	0.953	0.952	0.950	0.953	0.950	0.956
6	1.040	1.045	1.049	1.049	1.048	1.045	1.047	1.052	1.045	1.048
7	1.127	1.133	1.133	1.138	1.135	1.131	1.134	1.138	1.136	1.141
8	1.202	1.214	1.214	1.213	1.215	1.208	1.211	1.217	1.212	1.220
9	1.276	1.286	1.289	1.285	1.288	1.285	1.282	1.288	1.287	1.292
10	1.343	1.354	1.357	1.351	1.355	1.352	1.353	1.357	1.357	1.363
11	1.406	1.418	1.421	1.416	1.422	1.418	1.420	1.421	1.428	1.423
12	1.469	1.478	1.483	1.479	1.482	1.481	1.483	1.484	1.491	1.485
13	1.528	1.538	1.539	1.538	1.540	1.540	1.544	1.545	1.549	1.541
14	1.583	1.592	1.595	1.594	1.596	1.598	1.598	1.598	1.604	1.601
15	1.635	1.643	1.648	1.650	1.648	1.650	1.652	1.653	1.660	1.656

Table B.9 Simulated values of the HWM index at percentile 90 under A1* – A2*.
Notes: The underlying distribution consists of 10,000 independent HWM index values that are computed for samples that are drawn from a standard normal distribution.

n	3	4	5	6	7	8	9	10	50	100
K										
3	0.753	0.762	0.779	0.773	0.778	0.770	0.774	0.780	0.776	0.773
4	0.886	0.900	0.904	0.909	0.906	0.906	0.904	0.911	0.910	0.915
5	0.993	1.004	1.012	1.016	1.011	1.015	1.013	1.020	1.025	1.020
6	1.087	1.102	1.109	1.109	1.111	1.111	1.109	1.117	1.121	1.122
7	1.177	1.189	1.191	1.198	1.198	1.197	1.199	1.202	1.204	1.211
8	1.254	1.267	1.273	1.273	1.276	1.280	1.278	1.276	1.283	1.294
9	1.327	1.341	1.346	1.346	1.350	1.347	1.350	1.352	1.361	1.365
10	1.395	1.406	1.411	1.414	1.422	1.419	1.417	1.425	1.431	1.430
11	1.454	1.474	1.477	1.480	1.485	1.484	1.484	1.489	1.497	1.494
12	1.518	1.535	1.536	1.539	1.542	1.545	1.545	1.548	1.559	1.551
13	1.576	1.590	1.597	1.598	1.601	1.603	1.603	1.609	1.618	1.611
14	1.631	1.644	1.654	1.655	1.655	1.661	1.660	1.665	1.671	1.667
15	1.684	1.699	1.706	1.712	1.709	1.713	1.713	1.719	1.726	1.720

Table B.10 Simulated values of the HWM index at percentile 95 under A1*– A2*.
Notes: The underlying distribution consists of 10,000 independent HWM index values that are computed for samples that are drawn from a standard normal distribution.

n	3	4	5	6	7	8	9	10	50	100
K										
3	0.805	0.820	0.834	0.831	0.841	0.832	0.831	0.840	0.848	0.844
4	0.926	0.947	0.961	0.964	0.965	0.957	0.961	0.967	0.978	0.978
5	1.035	1.052	1.066	1.074	1.068	1.072	1.072	1.083	1.093	1.087
6	1.127	1.149	1.159	1.168	1.166	1.172	1.170	1.175	1.184	1.189
7	1.214	1.236	1.241	1.252	1.260	1.254	1.255	1.257	1.269	1.274
8	1.294	1.311	1.328	1.327	1.330	1.338	1.334	1.336	1.346	1.356
9	1.366	1.385	1.392	1.398	1.405	1.405	1.408	1.410	1.419	1.425
10	1.434	1.455	1.457	1.473	1.476	1.472	1.473	1.480	1.489	1.487
11	1.500	1.521	1.526	1.533	1.540	1.543	1.534	1.544	1.561	1.549
12	1.557	1.580	1.586	1.594	1.603	1.605	1.597	1.604	1.621	1.609
13	1.617	1.637	1.647	1.651	1.658	1.665	1.658	1.663	1.681	1.670
14	1.671	1.692	1.706	1.707	1.707	1.718	1.715	1.716	1.740	1.730
15	1.727	1.747	1.755	1.764	1.761	1.772	1.771	1.773	1.790	1.785

Table B.11 Simulated values of the HWM index at percentile 97.5 under A1*–A2*.
Notes: The underlying distribution consists of 10,000 independent HWM index values that are computed for samples that are drawn from a standard normal distribution.

n	3	4	5	6	7	8	9	10	50	100
K										
3	0.839	0.874	0.888	0.896	0.909	0.900	0.906	0.917	0.919	0.930
4	0.968	1.008	1.013	1.024	1.033	1.027	1.037	1.042	1.057	1.055
5	1.082	1.111	1.125	1.135	1.142	1.144	1.138	1.150	1.172	1.169
6	1.174	1.208	1.212	1.225	1.233	1.247	1.244	1.246	1.256	1.254
7	1.259	1.287	1.301	1.310	1.318	1.327	1.329	1.327	1.341	1.341
8	1.333	1.371	1.382	1.394	1.392	1.393	1.406	1.394	1.416	1.422
9	1.412	1.434	1.454	1.464	1.470	1.479	1.470	1.472	1.494	1.490
10	1.477	1.505	1.522	1.532	1.534	1.543	1.532	1.551	1.557	1.547
11	1.545	1.574	1.585	1.600	1.599	1.603	1.600	1.612	1.636	1.623
12	1.609	1.628	1.645	1.658	1.660	1.664	1.667	1.670	1.700	1.686
13	1.665	1.685	1.704	1.714	1.719	1.724	1.728	1.727	1.757	1.740
14	1.721	1.748	1.758	1.771	1.767	1.784	1.784	1.782	1.805	1.805
15	1.783	1.801	1.814	1.821	1.822	1.836	1.836	1.833	1.859	1.860

Table B.12 Simulated values of the HWM index at percentile 99 under A1* – A2*.
Notes: The underlying distribution consists of 10,000 independent HWM index values that are computed for samples that are drawn from a standard normal distribution.

REFERENCES

Anderson, T.W. and D.A. Darling (1952). Asymptotic theory of certain goodness of fit criteria based on stochastic processes. *Annals of Mathematical Statistics*, 23, 193-212.

Cramér, H. (1928). On the composition of elementary errors II: statistical applications. *Skandinavisk Aktuarietidskrift*, 11, 141-180.

Dufour, J.M. (1995). Monte Carlo tests with nuisance parameters: a general approach to finite-sample inference and non-standard asymptotics in econometrics. *Technical report C.R.D.E.*, Université de Montreal.

Dufour, J.M. and J.F. Kiviet (1998). Exact inference methods for first-order autoregressive distributed lags models. *Econometrica*, 66, 79-104.

Fisz, M. (1960). On a result by M. Rosenblatt concerning the von Mises-Smirnov test. *Annals of Mathematical Statistics*, 31, 427-429.

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

Hinloopen, J. (1997). *Research and development, product differentiation, and robust estimation.* Ph.D.-thesis, European University Institute.

Hinloopen, J. and R.J.L.M. Wagenvoort (2010). Identifying All Distinct Sample P-P Plots, with an Application to the exact Finite Sample Distribution of the L_1–FCvM Test Statistic. *Tinbergen Institute Discussion Paper*, TI 2010-083/1.

Holmgren, E.B. (1995). The p-p plot as a method of comparing treatment effects. *Journal of the American Statistical Society*, 90, 360-365.

IMF (2011). International Financial Statistics. http://www.imf.org (accessed March 31, 2012).

Johnson, B.McK. and T. Killeen (1983). An explicit formula for the C.D.F. of the L1 norm of the Brownian bridge. *The Annals of Probability*, 11, 807-808.

Kiefer, J. (1959). K-sample analogues of the Kolmogorov-Smirnov and Cramer-v.Mises tests. *Annals of Mathematical Statistics*, 30, 420-447.

Kolmogorov, A.N. (1933). Sulla determinizazione empirica delle leggi di probabilita. *Giornale dell 'Istituto Italiano Attuari*, 4, 1-11.

Kuiper, N.H. (1960). *Tests concerning random points on a circle*. Koninklijke Nederlandse Akademie van Wetenschappen, The Netherlands.

Rosenblatt, M. (1952). Limit theorems associated with variants of the von Mises statistic. *Annals of Mathematical Statistics*, 23, 617-623.

Scholz, F.W. and M.A. Stephens (1987). K-sample Anderson-Darling tests. *Journal of the American Statistical Association*, 82, 918-924.

Schmidt, F. and M. Trede (1995). A distribution free test for the two sample problem for general alternatives. *Computational Statistics & Data Analysis*, 20, 409-419.

Smirnov, N.V. (1939). On the deviation of the empirical distribution function. *Mathematicheskii Sbornik*, 6, 3-26.

Von Mises, R. (1931). *Wahrscheinlichkeitsrechnung*. Vienna: Deuticke.

Wilk, M.B. and R. Gnanadesikan (1968). Probability plotting methods for the analysis of data. *Biometrika*, 55, 1-17.