Chapter 6
How to Average Equating Functions, If You Must

Paul W. Holland and William E. Strawderman

6.1 Introduction and Notation

An interest in *averaging* two or more equating functions can arise in various settings. As the motivation for the *angle bisector* method described later in this paper, Angoff (1971) mentioned situations with multiple estimates of the same linear equating function for which averaging the different estimates may be appropriate. In the nonequivalent groups with anchor test (NEAT) equating design, several possible linear and nonlinear equating methods are available. These are based on different assumptions about the missing data in that design (von Davier, Holland, & Thayer, 2004b). It might be useful to average the results of some of the options for a final *compromise* method. Other recent proposals include averaging an estimated equating function with the *identity transformation* to achieve more stability in small samples (Kim, von Davier, & Haberman, 2008) as well as creating *hybrid* equating functions that are averages of linear and equipercentile equating functions, putting more weight on one than on the other (von Davier, Fournier-Zajac, & Holland, 2006). In his discussion of the angle bisector, Angoff implicitly weighted the two linear functions equally. The idea of *weighting* the two functions differently is a natural and potentially useful added flexibility to the averaging process that we use throughout our discussion.

We denote by $e_1(x)$ and $e_2(x)$ two different equating functions for linking scores on test X to scores on test Y. We will assume that $e_1(x)$ and $e_2(x)$ are *strictly increasing* continuous functions of x over the entire real line. The use of the entire real line is appropriate for both *linear* equating functions and for the method of

P.W. Holland (✉)
Paul Holland Consulting Corporation, 200 4th Ave South, Apt 100, St Petersburg, FL 33701, USA
e-mail: pholland@ets.org

W.E. Strawderman
Department of Statistics, Rutgers University, 110 Frelinghuysen Rd., Room 561 Hill Center Building for the Mathematical Sciences, Busch Campus, Piscataway, NJ N08854, USA
e-mail: straw@stat.rutgers.edu

A.A. von Davier (ed.), *Statistical Models for Test Equating, Scaling, and Linking*, Statistics for Social and Behavioral Sciences,
DOI 10.1007/978-0-387-98138-3_6,

kernel equating (von Davier et al., 2004b). Our main discussion concerns averages of equating functions that are defined for *all real* x.

Suppose it is desired to *average* $e_1(x)$ and $e_2(x)$ in some way, putting weight w on $e_1(x)$ and $1 - w$ on $e_2(x)$. In order to have a general notation for this, we will let $\oplus$ denote an *operator* that forms a weighted average of two such functions, e_1 and e_2, and puts weight w on e_1 and $1 - w$ on e_2. At this point we do not define exactly what $\oplus$ is and let it stand for any method of averaging. Our notation for any such weighted average of e_1 and e_2 is

$$we_1 \oplus (1-w)e_2 \tag{6.1}$$

to denote the resulting equating *function*. We denote its *value* at some X-score, x, by

$$we_1 \oplus (1-w)e_2(x). \tag{6.2}$$

If there are three such functions, e_1, e_2, e_3, then their *weighted average function* is denoted as

$$w_1e_1 \oplus w_2e_2 \oplus w_3e_3, \tag{6.3}$$

where the weights, w_i, sum to 1.

6.2 Some Desirable Properties of Averages of Equating Functions

Using our notation we can describe various properties that the operator, $\oplus$, should be expected to possess. The first five appear to be obvious requirements for any type of averaging process.

6.2.1 Property 1

Property 1: The order of averaging does not matter, so that

$$we_1 \oplus (1-w)e_2 = (1-w)e_2 \oplus we_1. \tag{6.4}$$

6.2.2 Property 2

Property 2: The weighted average should lie between the two functions being averaged, so that

$$\text{if } e_1(x) \le e_2(x), \text{then } e_1(x) \le we_1 \oplus (1-w)e_2(x) \le e_2(x). \tag{6.5}$$

Property 2 also implies the following natural property:

6.2.3 Property 3

Property 3: If the two equating functions are equal at a score, x, the weighted average has that same common value at x, so that

$$\text{if } e_1(x) = e_2(x), \text{then } we_1 \oplus (1-w)e_2(x) = e_1(x). \tag{6.6}$$

6.2.4 Property 4

It also seems reasonable for the average of two equating functions (that are always strictly increasing and continuous) to have both of these conditions as well. Thus, our next condition is Property 4: For any w, $we_1 \oplus (1-w)e_2(x)$ is a continuous and strictly increasing function of x.

6.2.5 Property 5

When it is desired to average three equating functions, as in Equation 6.3, it also seems natural to require the averaging process to get the same result as first averaging of a pair of the functions and then averaging that average with the remaining function, that is, Property 5: If w_1, w_2, w_3 are positive and sum to 1.0, then

$$w_1e_1 \oplus w_2e_2 \oplus w_3e_3 = w_1e_1 \oplus (1-w_1)[\frac{w_2}{1-w_1}e_2 \oplus \frac{w_3}{1-w_1}e_3]. \tag{6.7}$$

Again, without dwelling on notational issues in Equation 6.7, the order of the pair-wise averaging should not matter, either.

6.2.6 Property 6

There are other, less obvious assumptions that one might expect of an averaging operator for equating functions. One of them is Property 6: If e_1 and e_2 are *linear* functions then so is $we_1 \oplus (1-w)e_2$, for any w. We think that Property 6 is a reasonable restriction to add to the list, because one justification for the linear equating function is its simplicity. An averaging process that changed linear functions to a nonlinear one seems to us to add a complication where there was none before.

6.2.7 Property 7

In addition to Properties 1–6 for $\oplus$, there is one very special property that has long been regarded as important for any equating function—the property of *symmetry*. This means that linking X to Y via the function $y = e(x)$ is assumed to imply that the link from Y to X is given by the *inverse function*, $x = e^{-1}(y)$, as noted by Dorans and Holland (2000). The traditional interpretation of the symmetry condition when applied to averaging equating functions is that averaging the inverse functions, e_1^{-1} and e_2^{-1}, results in the *inverse* function of the *average* of e_1 and e_2.

Using our notation for $\oplus$, the *condition of symmetry* may be expressed as Property 7:

$$\text{For any } w,\ (w e_1 \oplus (1-w)e_2)^{-1} = we_1^{-1} \oplus (1-w)e_2^{-1}. \tag{6.8}$$

From Equation 6.8 we can see that the symmetry property requires that the averaging operator, $\oplus$, be formally *distributive* relative to the *inverse operator*.

6.3 The Point-Wise Weighted Average

The simplest type of weighted average that comes to mind is the simple *point-wise weighted average* of e_1 and e_2. It is defined as

$$m(x) = we_1(x) + (1-w)e_2(x), \tag{6.9}$$

where w is a fixed value, such as $w = 1/2$.

Geometrically, m is found by averaging the values of e_1 and e_2 along the vertical line located at x. For its heuristic value, our notation in Equations 6.1 and 6.2 was chosen to mimic Equation 6.9 as much as possible. In general, $m(x)$ in Equation 6.9 will satisfy Properties 1–6, for any choice of w. However, $m(x)$ will not always satisfy the symmetry property, Property 7. That is, if the inverses, $e_1^{-1}(y)$ and $e_2^{-1}(y)$, are averaged to obtain

$$m^*(y) = we_1^{-1}(y) + (1-\mathrm{w})e_2^{-1}(y), \tag{6.10}$$

then only in special circumstances will $m^*(y)$ be the inverse of $m(x)$ in Equation 6.9.

This is easiest to see when e_1 and e_2 are *linear*. For example, suppose e_1 and e_2 have the form

$$e_1(x) = a_1 + b_1x, \text{ and } e_2(x) = \ = a_2 + b_2x. \tag{6.11}$$

The point-wise weighted average of Equation 6.11 becomes

$$m(x) = \bar{a} + \bar{b}x, \tag{6.12}$$

where

$$\bar{b} = w\,b_1 + (1 - w)\,b_2 \;\; \text{and} \;\; \bar{a} = w\,a_1 + (1 - w)a_2. \tag{6.13}$$

However, the inverse functions for e_1 and e_2 are also linear with slopes $1/b_1$ and $1/b_2$, respectively. Thus, the point-wise average of the inverse functions, $m^*(x)$, has a slope that is the average of the reciprocals of the b_is:

$$b^* = w(1/b_1) + (1 - w)(1/b_2). \tag{6.14}$$

The inverse function of $m^*(x)$ is also linear and has slope $1/b^*$, where b^* is given in Equation 6.14. Thus, the slope of the inverse of $m^*(x)$ is the *harmonic* mean of b_1 and b_2. So, in order for the slope of the inverse of $m^*(x)$ to be the point-wise weighted average of the slopes of e_1 and e_2, the mean and the harmonic means of b_1 and b_2 must be equal. It is well known that this is only true if b_1 and b_2 are equal, in which case the equating functions are parallel. It is also easy to show that the intercepts do not add any new conditions. Thus we have Result 1 below.

6.3.1 Result 1

Result 1: The point-wise weighted average in Equation 6.9 satisfies the symmetry property for two linear equating functions if and only if the slopes, b_1 and b_2, are equal.

When $e_1(x)$ and $e_2(x)$ are non-linear they may still be *parallel* with a constant difference between them, that is

$$e_1(x) = e_2(x) + c \;\; \text{for all} \;\; x. \tag{6.15}$$

When Equation 6.15 holds, it is easy to establish Result 2.

6.3.2 Result 2

Result 2: If $e_1(x)$ and $e_2(x)$ are nonlinear but parallel so that Equation 6.15 holds, then the point-wise weighted average also will satisfy the symmetry property, Property 7. In this case, the point-wise average is simply a constant added (or subtracted) to either e_1 or e_2, for example,

$$m(x) = e_2(x) + wc = e_1(x) - (1 - w)c. \tag{6.16}$$

Thus, although the point-wise weighted average does not always satisfy the symmetry property, it does satisfy it if $e_1(x)$ and $e_2(x)$ are parallel curves or lines.

6.4 The Angle Bisector Method of Averaging Two Linear Functions

Angoff (1971) made passing reference to the *angle bisector* method of averaging two linear equating functions. In discussions with Angoff, Holland was informed that this method was explicitly proposed as a way of preserving the symmetry property, Property 7. Figure 6.1 illustrates the angle bisector, denoted by e_{AB}.

While the geometry of the angle bisector is easy to understand, for computations a formula is more useful. Holland and Strawderman (1989) give such a formula. We state their result next, and outline its proof in Section 6.5.

6.4.1 Result 3: Computation of the Unweighted Angle Bisector

Result 3: If $e_1(x)$ and $e_2(x)$ are two linear equating functions as in Equation 6.9 that intersect at a point, then the linear function that bisects the angle between them is the point-wise weighted average

$$e_{AB} = We_1 + (1 - W)e_2, \tag{6.17}$$

with W given by

$$W = \frac{(1 + b_1^2)^{-1/2}}{(1 + b_1^2)^{-1/2} + (1 + b_2^2)^{-1/2}}. \tag{6.18}$$

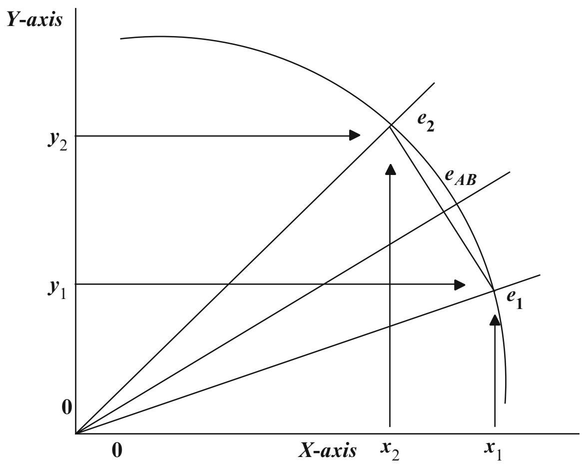

Fig. 6.1 The angle bisector is also the chord bisector

Note that in Result 3, if the two slopes are the same then $W = ½$ and the formula for the angle bisector reduces to the equally weighted point-wise average of the two parallel lines. It may be shown directly that the angle bisector given by Equations 6.17 and 6.18 satisfies the symmetry property, Property 7, for any two linear equating functions. Thus, in order for the point-wise weighted average Equation 6.9 to satisfy Property 7 for any pair of linear equating functions, it is necessary for the weight, w, to depend on the functions being averaged. It cannot be the same value for all pairs of functions.

6.5 Some Generalizations of the Angle Bisector Method

One way to understand the angle bisector for two linear functions is to imagine a circle of radius 1 centered at the point of intersection of the two lines. For simplicity, and without loss of generality, assume that the intersection point is at the origin, $(x, y) = (0, 0)$. This is also illustrated in Figure 6.1.

The linear function, e_i, intersects the circle at the point $(x_i, y_i) = (x_i, b_i x_i)$, and because the circle has radius 1 we have

$$\begin{aligned} &(x_i)^2 + (b_i x_i)^2 = 1, \text{ or} \\ &x_i = (1 + (b_i)^2)^{-1/2}. \end{aligned} \tag{6.19}$$

Thus, the linear function, e_i, intersects the circle at the point

$$(x_i, y_i) = ((1 + (b_i)^2)^{-1/2},\ b_i(1 + (b_i)^2)^{-1/2}). \tag{6.20}$$

The line that bisects the *angle* between e_1 and e_2 also bisects the *chord* that connects the intersection points, (x_1, y_1) and (x_2, y_2) given in Equation 6.20. The point of bisection of the chord is $((x_1 + x_2)/2, (y_1 + y_2)/2)$. From this it follows that the line through the origin that goes through the point of bisection of the chord has the slope,

$$b = \frac{y_1 + y_2}{x_1 + x_2} = Wb_1 + (1 - W)b_2, \tag{6.21}$$

where W is given by Equation 6.18. This shows that the angle bisector is the point-wise weighted average given in Result 3.

One way to generalize the angle bisector to include weights, as in Equation 6.1, is to divide the chord between (x_1, y_1) and (x_2, y_2) given in Equation 6.20 *proportionally* to w and $1 - w$ instead of *bisecting* it. If we do this, the point on the chord that is w of the way *from* (x_2, y_2) *to* (x_1, y_1) is $(wx_1 + (1-w)x_2, wy_1 + (1-w)y_2)$. It follows that the line through the origin that goes through this w-point on the chord has the slope

$$b = \frac{wy_1 + (1 - w)y_2}{wx_1 + (1 - w)x_2} = Wb_1 + (1 - W)b_2, \tag{6.22}$$

where W is now given by

$$W = \frac{w(1+b_1^2)^{-1/2}}{w(1+b_1^2)^{-1/2} + (1-w)(1+b_2^2)^{-1/2}}. \tag{6.23}$$

Hence, a weighted generalization of the angle bisector of two linear equating functions is given by Equation 6.17, with W specified by Equation 6.23.

This generalization of the angle bisector will divide the *angle* between e_1 and e_2 proportionally to w and $1 - w$ only when $w = ½$. Otherwise, this generalization only approximately divides the angle proportionately. In addition, direct calculations show that this generalization of the angle bisector will satisfy all of the properties, Properties 1–7.

However, the angle bisector may be generalized in other ways as well. For example, instead of a circle centered at the point of intersection, suppose we place an L_p-circle there instead. An L_p-circle is defined by Equation 6.24:

$$|x|^p + |y|^p = 1, \tag{6.24}$$

where $p > 0$. Examples of L_p-circles for various choices of $p = 1$ and 3 are given in Figures 6.2 and 6.3.

If we now use the chord that connects the intersection points of the two lines with a given L_p-circle, as we did above for the ordinary circle, we find the following

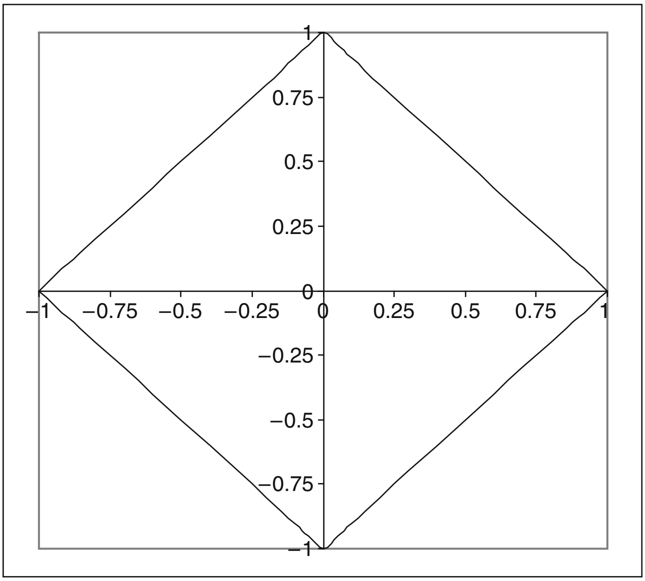

Fig. 6.2 Plot of the unit L_p-circle, $p = 1$

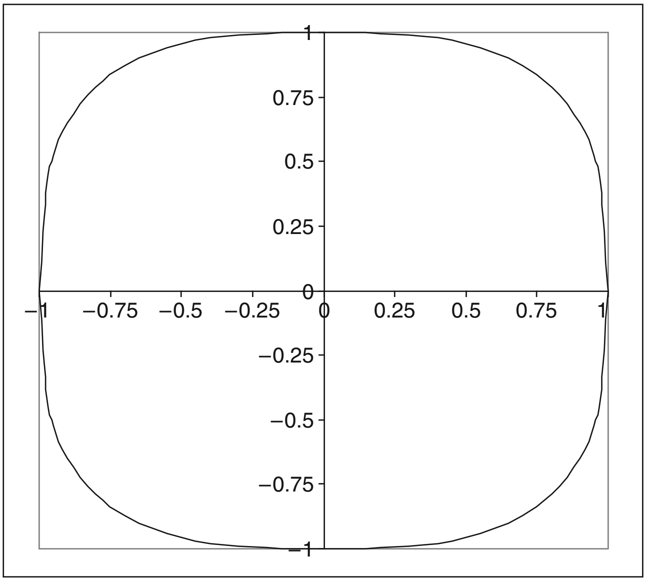

Fig. 6.3 Plot of the unit L_p-circle, $p = 3$

generalization of the angle bisector. We form the point-wise weighted average in Equation 6.17, but we use as W the following weight:

$$W = \frac{w(1 + b_1^p)^{-1/p}}{w(1 + b_1^p)^{-1/p} + (1 - w)(1 + b_2^p)^{-1/p}}, \tag{6.25}$$

for some $p > 0$, and $0 < w < 1$. It is a simple exercise to show that the use of W from Equation 6.25 as the weight in Equation 6.17 also will satisfy Properties 1–7 for any choice of $p > 0$, and $0 < w < 1$. We will find the case of $p = 1$ of special interest later. In that case W has the form

$$W = \frac{w(1 + b_1)^{-1}}{w(1 + b_1)^{-1} + (1 - w)(1 + b_2)^{-1}}. \tag{6.26}$$

Thus, the system of weighted averages (Equation 6.17) with weights that depend on the two slopes, as in Equation 6.25, produces a variety of ways to average linear equating functions that all satisfy Properties 1–7. Thus, the angle bisector is seen to be only one of an infinite family of possibilities. It is worth mentioning here that when $w = ½$, all of these averages of two linear equating functions using Equation 6.25 have the property of putting *more* weight on the line with the *smaller* slope. As a simple example, if $b_1 = 1$ and $b_2 = 2$, then for w = ½, Equation 6.26 gives the value $W = 0.6$ for the case of $p = 1$.

An apparent limitation of all of these circle methods of averaging two linear functions is that they do not immediately generalize to the case of three or more such functions. When there are three functions, they do not necessarily meet at a point; there could be three intersection points. In such a case, the idea of using an L_p-circle centered at the "point of intersection" makes little sense. However, the condition Property 5 gives us a way out of this narrow consideration. Applying it to the point-wise weighted average results obtained so far, it is tedious but straightforward to show that the *multiple function generalization* of Equation 6.17 coupled with Equation 6.25 is given by Result 4.

6.5.1 Result 4

Result 4: If $\{w_i\}$ are positive and sum to 1.0 and if $\{e_i\}$ are linear equating functions, then Property 5 requires that the pair-wise averages based on Equations 6.17 and 6.25 lead to

$$w_1e_1 \oplus w_2e_2 \oplus w_3e_3 \oplus \ldots = \sum_i W_ie_i \tag{6.27}$$

where

$$W_i \frac{w_i(1+b_i^p)^{-1/p}}{\sum_j w_j(1+b_j^p)^{-1/p}}. \tag{6.28}$$

Result 4 gives a solution to the problem of averaging several different linear equating functions that is easily applied in practice, once choices for p and w are made.

Holland and Strawderman (1989) introduced the idea of the *symmetric weighted average* (swave) of two equating functions that satisfies conditions of Properties 1–7 for any pair of linear or nonlinear equating functions. In the next two sections we develop a generalization of the symmetric average.

6.6 The Geometry of Inverse Functions and Related Matters

To begin, it is useful to illustrate the geometry of a strictly increasing continuous function, $y = e(x)$, and its inverse, $x = e^{-1}(y)$. First, fix a value of x in the domain of $e(\cdot)$, and let $y = e(x)$. Then the four points, (x, y), (x, x), (y, y) and (y, x), form the four corners of a square in the (x, y) plane, where the length of each side is $|x - y|$. The two points, (x, x) and (y, y), both lie on the 45-degree line; the other two points lie on opposite sides of the 45-degree line on a line that is at right angles, or orthogonal, to it. In addition, (x, y) and (y, x) are equidistant from the 45-degree line. However, by definition of the inverse function, when $y = e(x)$, it is also the case that $x = e^{-1}(y)$. Hence, the four points mentioned above can be re-expressed as $(x, e(x))$, (x, x), $(e(x), e(x))$, and $(y, e^{-1}(y))$, respectively.

The points $(x, e(x))$ and $(y, e^{-1}(y))$ are equidistant from the 45-degree line and on opposite sides of it. Furthermore, the line connecting them is orthogonal to the 45-degree line and is bisected by it. These simple facts are important for the rest of this discussion. For example, from them we immediately can conclude that the graphs of $e(\cdot)$ and $e^{-1}(\cdot)$ are reflections of each other about the 45-degree line in the (x, y) plane. This observation is the basis for the swave defined in Section 6.7.

Another simple fact that we will make repeated use of is that a strictly increasing continuous function of x, $e(x)$, crosses any line that is orthogonal to the 45-degree line in exactly one place. This is illustrated in Figure 6.4 for the graphs of two functions. In order to have a shorthand term for lines that are orthogonal to the 45-degree line, we will call them the *orthogonal lines* when this is unambiguous.

We recall the elementary fact that the equation for what we are calling an orthogonal line is

$$y = -x + c, \text{ or } y + x = c, \text{ for some constant, } c. \tag{6.29}$$

Thus, we have the relationship

$$e(x_1) + x_1 = c = y + x \tag{6.30}$$

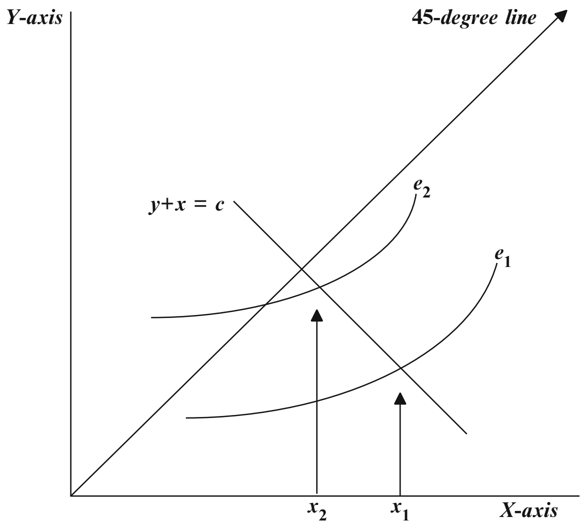

Fig. 6.4 An illustration of the intersections of $e_1(x)$ and $e_2(x)$ with an orthogonal line

for any other point, (x, y), that is on the orthogonal line. Equation 6.30 plays an important role in the definition of the *swave* in Section 6.7. Finally, we note that if e is a strictly increasing continuous function, then its inverse, e^{-1}, is one as well.

6.7 The Swave: The Symmetric *w*-Average of Two Equating Functions

With this preparation, we are ready to define the symmetric w-average or swave of two linear or nonlinear equating functions, $e_1(x)$ and $e_2(x)$. Note that from the above discussion, any orthogonal line, of the form given by Equation 6.29, will intersect $e_1(x)$ at a point, x_1, and $e_2(x)$ at another point, x_2. This is also illustrated in Figure 6.4.

The idea is that the value of the swave, $e_w(\cdot)$, is given by the point on the orthogonal line that corresponds to the weighted average of the two points, $(x_1, e_1(x_1))$ and $(x_2, e_2(x_2))$:

$$(\bar{x}, e_w(\bar{x})) = \mathrm{w}\big(x_{1,}e_1(x_1)\big) + (1 - \mathrm{w})(x_2, e_2(x_2)). \tag{6.31}$$

This is illustrated in Figure 6.5.

The point, $((\bar{x}, e_w(\bar{x})))$, is the weighted average of the two points, $(x_1, e_1(x_1))$ and $(x_2, e_2(x_2))$. Thus,

$$\bar{x} = wx_1 + (1 - w)x_2, \tag{6.32}$$

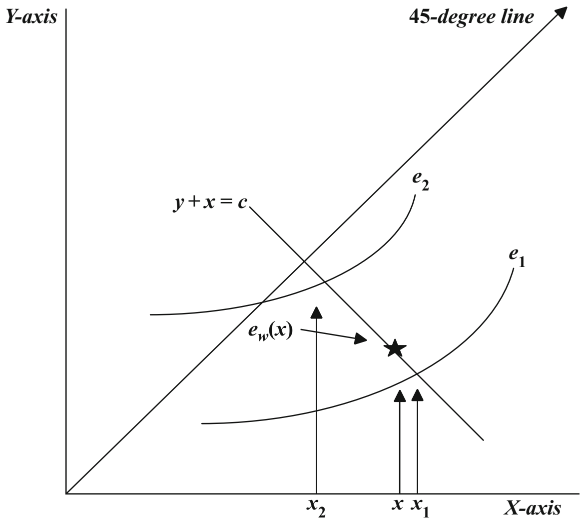

Fig. 6.5 An illustration of the swave of $e_1(x)$ and $e_2(x)$ at $x = w\,x_1 + (1 - w\,)\,x_2$ for a w greater than ½

and

$$e_w(\bar{x}) = we_1(x_1) + (1 - w)e_2(x_2). \tag{6.33}$$

In Equation 6.31, $\bar{x}$ is given by Equation 6.32. In order to define $e_w(x)$ for *an arbitrary* point, x, we start with x and define $x_1 = x - (1 - w)t$ and $x_2 = x + wt$ for some, as yet unknown, positive or negative value, t. Note that from the definitions of x_1 and x_2, their weighted average, $wx_1 + (1 - w)x_2$, equals x, so the given x can play the role $\bar{x}$ of Equation 6.32.

Next, we find a value of t such that $(x_1, e_1(x_1))$ and $(x_2, e_2(x_2))$ lie on the same orthogonal line, as in Figure 6.5. From Equation 6.30, this condition on t requires that Equation 6.34 is satisfied:

$$e_1(x_1) + x_1 = e_2(x_2) + x_2. \tag{6.34}$$

Equation 6.34 may be expressed in terms of x and t as

$$e_1(x - (1 - w)t) + x - (1 - w)t = e_2(x + wt) + x + wt.$$

or

$$t = e_1(x - (1 - w)t) - e_2(x + wt). \tag{6.35}$$

Equation 6.35 plays an important role in what follows.

In general, for each value of x, Equation 6.35 is a nonlinear equation in t. As we show in the Appendix, for any choice of x and w and for any strictly increasing continuous equating functions, e_1 and e_2, Equation 6.35 always has a *unique* solution for t. The solution of Equation 6.35 for t implicitly defines t as a function of x, which we denote by $t(x)$. Once $t(x)$ is in hand, the value of the swave at x, $e_w(x)$, is computed from the expression,

$$e_w(x) = we_1(x - (1 - w)t(x)) + (1 - w)e_2(x + wt(x)). \tag{6.36}$$

The definition of e_w in Equation 6.36 is an example of the operator $\oplus$ in Equation 6.1. In Equation 6.36, there is a clear sense in which the weight w is applied to e_1 and $1 - w$ is applied to e_2. We show later that the swave differs from the point-wise weighted average in Equation 6.9, except when the two equating functions are parallel, as discussed above. Moreover, the definition of the swave is a *process* that requires the whole functions, e_1 and e_2, rather than just their evaluation at the selected x-value. In the Appendix we show that the solution for t in Equations 6.35 and 6.36 is unique.

In the Appendix we show that the swave satisfies conditions in Properties 2 and 4. We discuss below the application of the swave to linear equating functions and show that it satisfies Property 6. That the swave satisfies Property 7, the symmetry property, is given in Result 5, next.

6.7.1 Result 5

Result 5: The swave, $e_w(x)$, defined by Equations 6.35 and 6.36, satisfies the symmetry property, Property 7.

Proof. Suppose we start with the inverse functions, e_1^{-1} and e_2^{-1}, and form their swave, denoted $e*_w(y)$, for a given y-value. Then Equations 6.35 and 6.36 imply that for any choice of y there is a value $t*$ that satisfies

$$t^* = e_1^{-1}(y - (1 - w)t^*) - e_2^{-1}(x + wt^*) \tag{6.37}$$

and

$$e^*{}_w(y) = we_1^{-1}(y - (1 - w)t^*) + (1 - w)e_2^{-1}(y + wt^*). \tag{6.38}$$

Now let

$$y_1 = y - (1 - w)t^*, \text{and } y_2 = y + wt^*.$$

Also, define x, x_1, and x_2 by

$$x = e^*{}_w(y), x_1 = e_1^{-1}(y_1), \text{ and } x_2 = e_2^{-1}(y_2). \tag{6.39}$$

Hence, by definition of the inverse,

$$y_1 = e_1(x_1), y_2 = e_2(x_2), \text{and } y = e^{*-1}_w(x). \tag{6.40}$$

From the definition of the swave, the following three points are all on the same orthogonal line:

$$(y, e^*{}_w(y)), (y_1, e_1^{-1}(y_1)), \text{and } (y_2, e_2^{-1}(y_2)).$$

However, using the relationships in Equations 6.38 and 6.39, these three points are the same as the following three points, which are also on that orthogonal line:

$$\left(e^*{}_w{}^{-1}(x), x\right), (e_1(x_1), x_1), \text{ and } (e_2(x_2), x_2).$$

Furthermore, the following three points are on that same orthogonal line:

$$\left(x, e^*{}_w{}^{-1}(x)\right), (x_1, e_1(x_1)), \text{and } (x_2, e_2(x_1)).$$

Yet, from Equation 6.38 it follows that

$$x = wx_1 + (1 - w)x_2, \tag{6.41}$$

so we let $t = x_2 - x_1$, and therefore, $x_1 = x - (1 - w)t$, and $x_2 = x + wt$.

Furthermore, from the definitions of y_1 and y_2, we have

$$y = wy_1 + (1 - w)y_2$$

and therefore

$$y = we_1(x_1) + (1 - w)e_2(x_2),$$

so that

$$y = e^{*-1}_{w}(x) = we_1(x - (1 - w)t) + (1 - w)e_2(x + wt). \tag{6.42}$$

Thus, the inverse function, $e{*_w}^{-1}$, satisfies Equation 6.36 for the swave. The only question remaining is whether the value of t in Equation 6.42 satisfies Equation 6.35.

However, the points $(x_1, e_1(x_1))$ and $(x_2, e_2(x_2))$ are on the same orthogonal line. Therefore, they satisfy Equation 6.34, from which Equation 6.35 for t follows.

This shows that the inverse function, $e{*_w}^{-1}$, satisfies the condition of the symmetric w-average, e_w, so that from the uniqueness of the solution to Equation 6.35 we have $e{*_w}^{-1} = e_w$, which proves Result 5. From the definitions of y_1, y_2, x_1, and x_2 in the proof of Result 5, it is easy to see that the t^* that solves Equation 6.37 for the inverse functions and the t that solves Equation 6.35 for the original functions are related by $t^* = x_1 - x_2 = -t$, so that t and t^* have the same *magnitude* but the opposite *sign*.

6.8 The Swave of Two Linear Equating Functions

In this section we examine the form of the swave in the linear case. The equation for t, Equation 6.35, now becomes linear in t and can be solved explicitly. So assume that

$$e_1(x) = a_1 + b_1x, \text{and}\, e_2(x) = a_2 + b_2x. \tag{6.43}$$

Then, Equation 6.35 is

$$t = a_1 + b_1(x - (1 - w)t) - a_2 - b_2(x + wt).$$

Hence,

$$t(1 + (1 - w)b_1 + wb_2) = (a_1 - a_2) + (b_1 - b_2)x$$

so that

$$t(x) = \frac{(a_1 - a_2) + (b_1 - b_2)x}{1 + (1 - w)b_1 + wb_2}. \tag{6.44}$$

Substituting the value of $t(x)$ from Equation 6.44 into the equation for $e_w(x)$ in Equation 6.36 results in

$$e_w(x) = \bar{a} + \bar{b}x - w(1-w)(b_1 - b_2)\frac{a_1 - a_2 + (b_1 - b_2)x}{1 + (1-w)b_1 + wb_2}, \tag{6.45}$$

where $\bar{a} = wa_1 + (1-w)a_2$ and $\bar{b} = wb_1 + (1-w)b_2$ denote the weighted averages of the intercepts and slopes of e_1 and e_2, respectively.

From Equation 6.45 we immediately see that, in the linear case, the swave, $e_w(x)$, is identical to the point-wise weighted average in Equation 6.9 if and only if the two slopes, b_1 and b_2, are identical, and the two linear functions are parallel. Simplifying Equation 6.45 further we obtain

$$e_w(x) = We_1(x) + (1-W)e_2(x), \tag{6.46}$$

where

$$W = \frac{w(1+b_1)^{-1}}{w(1+b_1)^{-1} + (1-w)(1+b_2)^{-1}}. \tag{6.47}$$

Thus, in the linear case, the swave is exactly the point-wise weighted average that arises for an L_p-circle with $p = 1$, in other words, Equation 6.26, discussed in Section 6.5. From Result 5, we know that the swave always satisfies the symmetry condition, Property 7, but this is also easily shown directly. We see that, in the linear case, the swave also satisfies Property 6.

Chapter 6 Appendix

6.A.1 Computing the Swave for Two Equating Functions

The key to computing e_w is Equation 6.35. This equation for $t(x)$ is nonlinear in general, so computing $t(x)$ requires numerical methods. A derivative-free approach that is useful in this situation is *Brent's method*. To use this method to solve Equation 6.35 for t we first define $g(t)$ as follows:

$$g(t) = t - e_1(x - (1-w)t) + e_2(x + wt). \tag{6.A.1}$$

If t_0 solves Equation 6.35, then t_0 is a zero of $g(t)$ in Equation 6.A.1. Brent's method is a way of finding the zeros of functions. It requires that two values of t are known, one for which $g(t)$ is positive and one for which $g(t)$ is negative. Theorem 1 summarizes several useful facts about $g(t)$ and provides the two needed values of t for use in Brent's method.

Theorem 1. *If e_1 and e_2 are strictly increasing continuous functions, then $g(t)$ defined in Equation 6.A.1 is a strictly increasing continuous function that has a unique zero at t_0. Furthermore, t_0 is positive if and only if $e_1(x) - e_2(x)$ is positive. Consequently, if $e_1(x) - e_2(x)$ is positive, then $g(0)$ is negative and $g(e_1(x) - e_2(x))$ is positive; furthermore, if $e_1(x) - e_2(x)$ is negative, then $g(0)$ is positive and $g(e_1(x) - e_2(x))$ is negative.*

Proof. The functions t, $-e_1(x-(1-w)\,t)$, and $e_2(x+w\,t)$ are all strictly increasing continuous functions of t so that their sum, $g(t)$, is also a strictly increasing continuous function of t. Hence, if $g(t)$ has a zero at t_0, this is its only zero. In order to show that $g(t)$ does have a zero at some t_0 it suffices to show that, for large enough t, $g(t) > 0$ and, for small enough t, $g(t) < 0$. But if $t > 0$, it follows from the strictly increasing (in t) nature of $-e_1(x-(1-w)\,t)$ and of $e_2(x+w\,t)$ that

$$g(t) > t - [e_1(x) - e_2(x)]. \tag{6.A.2}$$

The right side of Equation 6.A.2 is greater than 0 if t is larger than $e_1(x) - e_2(x)$. Similarly, if $t < 0$, it also follows that

$$g(t) < t - [e_1(x) - e_2(x)]. \tag{6.A.3}$$

The right side of Equation 6.A.3 is less than 0 if t is less than $e_1(x) - e_2(x)$. Hence, these two inequalities show that $g(t)$ always has a single zero at a value we denote by t_0.

Now, suppose that $t_0 > 0$. Then $g(t_0) = 0$ by definition so that

$$0 < t_0 = e_1(x-(1-w)\,t_0) - e_2(x+w\,t_0) \tag{6.A.4}$$

But by the strict monotonicity of e_1 and e_2, we have

$$e_1(x-(1-w)\,t_0) < e_1(x), \text{and} - e_2(x+w\,t_0) < -e_2(x)$$

so that

$$e_1(x-(1-w)\,t_0) - e_2(x+wt_0) < e_1(x) - e_2(x). \tag{6.A.5}$$

Combining Equations 6.A.4 and 6.A.5 shows that if $t_0 > 0$, then $e_1(x) - e_2(x) > 0$. A similar argument shows that if $t_0 < 0$, then $e_1(x) - e_2(x) < 0$. Hence t_0 is positive if and only if $e_1(x) - e_2(x)$ is positive. Note that we can always compute $e_1(x) - e_2(x)$ because it is assumed that these functions are given to us. Thus, from the relative sizes of $e_1(x)$ and $e_2(x)$ we can determine the sign of the zero, t_0.

Because $g(t)$ is strictly increasing we have the following additional result. If $e_1(x) - e_2(x)$ is positive, then t_0 is also positive and therefore $g(0)$ is negative. Also, if $e_1(x) - e_2(x)$ is negative, then t_0 is also negative and therefore $g(0)$ is positive.

Now suppose again that $e_1(x) - e_2(x)$ is positive so that t_0 is also positive. However, from Equation 6.19, for any positive t, $g(t) > t - [e_1(x) - e_2(x)]$, so let $t = t_0$. Hence,

$$0 = g(t_0) > t_0 - [e_1(x) - e_2(x)], \tag{6.A.6}$$

so that

$$0 < t_0 < e_1(x) - e_2(x). \tag{6.A.7}$$

Hence, g($e_1(x) - e_2(x)$) is positive as well. Thus, whenever $e_1(x) - e_2(x)$ is positive, then g(0) is negative and g($e_1(x) - e_2(x)$) is positive. When $e_1(x) - e_2(x)$ is negative, a similar argument shows that

$$e_1(x) - e_2(x) < t_0 < 0. \tag{6.A.8}$$

Hence g($e_1(x) - e_2(x)$) is negative. This finishes the proof of Theorem 1.

6.A.2 *Properties of the Swave*

Theorem 2. *The swave, $e_w(x)$, satisfies Property* 2 *and lies strictly between $e_1(x)$ and $e_2(x)$, for all x.*

Proof. Consider the case when $e_1(x) > e_2(x)$ (the reverse case is proved in a similar way). We wish to show that $e_1(x) > e_w(x) > e_2(x)$. Because $e_1(x) > e_2(x)$, from Theorem 1 it follows that $t(x) > 0$ as well. From the strictly increasing natures of e_1 and e_2, it follows that

$$e_1(x_1) < e_1(x), \text{and}\, e_2(x_2) > e_2(x).$$

We wish to show that $e_1(x) > e_w(x) > e_2(x)$, so consider first the upper bound. By definition,

$$e_w(x) = we_1(x_1) + (1 - w)e_2(x_2) < we_1(x) + (1 - w)e_2(x_2).$$

However,

$$0 < t(x) = e_1(x_1) - e_2(x_2), \text{so that } e_2(x_2) < e_1(x_1) < e_1(x).$$

Combining these results give us

$$e_w(x) < we_1(\mathrm{x}) + (1 - w)e_1(x) = e_1(x),$$

the result we wanted to prove. The lower bound is found in an analogous manner.

Theorem 3. *The swave is strictly increasing if* e_1 *and* e_2 *are.*

Facts: $e(x)$ monotone implies $c(x) = x + e(x)$ is strictly monotone (since it is a sum of a monotone and a strictly monotone function). Also, $c(x^*) > c(x)$ implies $x^* > x$ and $e(x^*) > e(x)$.

Let $c_i(x) = x + e_i(x)$, $\mathrm{i} = 1, 2$. Also let $e_w(x) = we_1(x_1) + (1-w)e_2(x_2)$, where $x = wx_1 + (1-w)x_2$ and $c_1(x_1) = c_2(x_2)$, i.e., $x_1 + e_1(x_1) = x_2 + e_2(x_2)$ so that $(x_i, e_i(x_i))$ are on same orthogonal line.

Assume $e_i(x)$ are both monotone increasing. Now suppose $x^* > x$ where $x = wx_1 + (1-w)x_2$ and $x^* = wx_1{}^* + (1-w)x_2{}^*$ and suppose further that $c_1(x_1) = c_2(x_2)$ and that $c_1(x_1{}^*) = c_2(x_2{}^*)$. Then, $(x_i, e_i(x_\mathrm{i}))$ are both on the same orthogonal line and $(x_i{}^*, e_i(x_i{}^*))$ are too (but possibly a different line). We want to conclude that $x_1{}^* > x_1$ and $x_2{}^* > x_2$. This will allow us to conclude that $e_i(x_i{}^*) > e_i(x_i)$ and hence that $e_w(x^*) > e_w(x)$, thereby proving the monotonicity of e_w.

Proof. Assume to the contrary that $x_1{}^* \leq x_1$. Then $c_2(x_2{}^*) = c_1(x_1{}^*) \leq c_1(x_1) = c_2(x_2)$, so that $x_2{}^* \leq x_2$. This in turn implies that $x^* = wx_1{}^* + (1-w)x_2{}^* \leq wx_1 + (1-w)x_2 = x$, or $x^* \leq x$, contradicting the assumption that $x^* > x$. A similar argument shows that $x_2{}^* > x_2$. Hence, $e_i(x_i{}^*) > e_i(x_i)$ and $e_w(x^*) > e_w(x)$, thereby proving the monotonicity of e_w.