Multi-dimensional Error Measures for Ex Post Facto Forecast and Estimation Evaluations[1]

David A. Swanson
Department of Sociology & Blakely Center for Sustainable Suburban Development
University of California Riverside
Riverside, CA 92521 USA
(David.swanson@ucr.edu)

Abstract

This paper deals with an issue that appears to be unexplored by demographers. The issue Is based on the question: "if one was not only doing ex post facto evaluations of estimates (or projections) of the total population but by characteristics such as race and geography, how would one summarize the summary measures of error?" The paper examines this issue in two dimensions using a standard summary measure of ex post facto error "Mean Algebraic Percent Error" (MALPE). It offers hypothetical, but empirical, examples of the issue using MALPE. It is aimed at answering the underlying question, which is "what is the nature of the relationship, if any, of summary errors of measure taken across different dimensions in conducting an ex post facto evaluation of the accuracy of a set of estimates or projections. The paper finds that relationships do exist and offers proof of these relationships both mathematically and empirically. It concludes by noting work to be done includes generalizing these findings to more than two dimensions and extending the findings to other commonly used summary measures of ex post facto error, such as Mean Absolute Percent Error (MAPE) and Root Mean Square Error (RMSE).

I. The Problem

Suppose we have done estimates back in 2009 for the year 2010 by race (using three race groups) for a state that has four counties. We now want to do an ex post facto evaluation of the accuracy of our estimates. Key to this evaluation is the summary measures of error for (1) Race, (2) Geography, and (3) for the race-geography jointly (i.e., the "overall MALPE"). As you will see below using the example of Mean Algebraic Percent Error (MALPE), the MALPE for race is not the same as the MALPE for geography and neither are the same as the MALPE for race-geography jointly. In table 1 below are the estimates (Eij) and in Table 2 are the actual populations (Pij). Table 3 contains the errors and Table 4 contains the proportionate errors (Eij-Pij)/Pij)).

Note that in each case, we have a 3 (race groups) by 4 (geography) table, with marginal totals for race in the fifth row and marginal totals for geography in the fourth column.

Table 1. 2010 Estimates (Eij) for the state by race and county

	estimate race1	race2	race3	total
geog1	305	100	55	460
geog2	500	125	40	665
geog3	86	10	5	101
geog4	125	20	10	155
total	1016	255	110	1381

Table 2. 2010 Actual Population (Pij) Counts for the state by race and county

	race1	race2	race3	total
geog1	291	98	45	434
geog2	490	130	51	671
geog3	85	8	4	97
geog4	130	25	12	167
total	996	261	112	1369

Table 3. Numerical Error (Eij-Pij) in the 2010 Estimates by race and county

	race1	race2	race3	total
geog1	14	2	10	26
geog2	10	-5	-11	-6
geog3	1	2	1	4
geog4	-5	-5	-2	-12
total	20	-6	-2	12

Table 4. Proportionate Error ((Eij-Pij)/Pij) in 2010 Estimates by race and county

	race1	race2	race3	total
geog1	0.04811	0.020408	0.222222	0.059908
geog2	0.020408	-0.03846	-0.21569	-0.00894
geog3	0.011765	0.25	0.25	0.041237
geog4	-0.03846	-0.2	-0.16667	-0.07186
total	0.02008	-0.02299	-0.01786	0.008766

Using the Preceding data, here are the MALPEs for Race, Geography, and jointly by race-geography, respectively.

Table 5. MALPE by Dimension

Dimension MALPE

RACE	-0.0069
GEOG	0.0051
JOINT	0.0136

In percent terms we have Race MALPE = -.69%, Geog MALPE = .51% and joint MALPE = 1.36%

As you can see, not only is it the case that none of the MALPEs is equal to another, but there is no clear relationship among them.

Now imagine yourself in front of a group of "stakeholder" discussing these estimates and you are describing these MALPEs (and how accurate your estimates are) when you are asked "why are they different?" The answer is not apparent, likely because it appears that demographers have not yet addressed the issue of simultaneous measures of error on multiple dimensions of a forecast or set of estimates. What follows is an initial attempt to answer the question, "Why are they different?" As you will see, the answer I believe is found in a subtle way that has to do with weighting.

II. MALPE and Weighted MALPE (WMALPE)

First recall, from Table 5

Dimension	MALPE
RACE	-0.0069
GEOG	0.0051
JOINT	0.0136

Now compare each of the (unweighted) MALPEs from Table 5 with its respective counterpart, WMALPE, as shown in Table 6.

Table 6. WMALPE by Dimension

RACE	0.0088
GEOG	0.0088
JOINT	0.0088

The obvious difference is that for each dimension, WMALPE is the same, which I believe is the key to understanding the differences observed in Table 5.

First, let's look at some definitions.

Let Ei = estimate (or projection) for population in category i (e.g., race, geography, etc.)
Pi = actual population in category i (e.g., race, geography, etc.)
k = number of categories

In general terms, MALPE and WMALPE can be expressed as follows:

$$MALPE = (1/k)*[\sum ((Ei-Pi)/Pi)] \quad [1]$$

And since $(1/k) = [1/(\sum Pi/k)/\sum Pi]$ [1.a]

$$MALPE = [1/(\sum Pi/k)/\sum Pi]^* [\sum ((Ei-Pi)/Pi)] \quad [1.b]$$

$$= /(\sum(Pi/k)/\sum Pi]^*[\sum ((Ei-Pi)/Pi)] \quad [1.c]$$

$$= [\sum((Ei-Pi)/(k^*Pi))] \quad [1.d]$$

$$= [\sum((Ei)/(k^*Pi))] -1 \quad [1.e]$$

$$WMALPE = \sum [(Pi/\sum Pi)^*((Ei-Pi)/Pi)] \quad [2]$$

And since $WMALPE = \sum [(Pi/\sum Pi)^*((Ei-Pi)/Pi)]$ [2.a]

$$= \sum[Ei-Pi)/\sum Pi] \quad [2.b]$$

$$= \sum Ei/\sum Pi - \sum Pi/\sum Pi \quad [2.c]$$

$$= [\sum Ei/\sum Pi] - 1 \quad [2.d]$$

Thus, we have

$$MALPE = [\sum((Ei)/(k^*Pi))] -1 \quad [3]$$

and $1 + MALPE = [\sum((Ei)/(k^*Pi))]$ [3.a]

$$WMALPE = [(\sum Ei)/(\sum Pi)] -1 \quad [4]$$

and $1 + WMALPE = [(\sum Ei)/(\sum Pi)]$ [4.a]

In other words:

MALPE can be found by subtracting 1.00 from the sum of the ratios of: the Estimated Population in category i to the Actual Population in category i times the number of categories (k).

1 + MALPE is the sum of the ratios of: the Estimated Population in category i to the Actual Population in category i times the number of categories (k).

WMALPE can be found by subtracting 1.00 from the sum of the estimated population divided by the sum of the actual population.

1+ WMALPE is the sum of the estimated population divided by the sum of the actual population.

NOTE: The re-expressions found in [3] and [4] represent short-cut ways to calculate MALPE and WMALPE, respectively.

Using the re-expressions in [3] and [4] the following relationship can be seen between WMALPE and MALPE. The proof is found in Appendix A.

$$WMALPE = MALPE + [(\sum Ei)/(\sum Pi)] - [\sum(Ei/kPi)] \qquad [5]$$

$$MALPE = WMALPE - [(\sum Ei)/(\sum Pi)] + [\sum(Ei/kPi)] \qquad [6]$$

The proof for [5] is found in Appendix A. Since [6] follows naturally from [5], no proof for it is needed, although the relationship shown in [6] is repeated in the Appendix following the proof for [5].

WMALPE can be viewed as MALPE + the sum of the estimated Population divided by the sum of actual population from which is subtracted the sum of the (1 = 1 to k) ratios of estimated populations to the actual populations with the latter weighted by a constant, k

Although I have not yet done a proof, it appears clear from [4] that WMALPEs for two dimensions (e.g., race and geography) will not only equal one another but also the joint WMALPE. This extends to n dimensions, such that all n WMALPEs = one another and their joint WMALPE. It appears that we also have general statements ([5] and [6]) for the relationship between a MALPE on a given dimension and its equivalent WMALPE. With this knowledge in hand, I know turn to specifying the relationships between a set of dimensional MALPEs, their joint MALPE and the corresponding WMALPE.

To be more specific, again using the two example dimensions of race and geography (but noting that this results can be generalized to more than 3 columns and 4 rows), we have

$$RMALPE = WMALPE - [(\sum Ei)/(\sum Pi)] + [\sum(Ei/k*Pi)] \qquad [7]$$

$$GMALPE = WMALPE - [(\sum Ej)/(\sum Pj)] + [\sum(Ej/(L*Pj))] \qquad [8]$$

and

$$JMALPE = WMALPE - [(\sum Eij)/(\sum Pij)] + [\sum(Eij/(m*Pij))] \qquad [9]$$

The Proof for Equation [9] is found in Appendix B.

where k = number of columns
L = number of rows
M = number of cells (k*L)

Using the preceding equations, we have from our example data in tables 1 and 2, the respective values for RMALPE, GMALPE, and JMALPE:

RMALPE=	WMALPE -	[(∑Ei)/(∑Pi)]	+	[∑(Ei/(k*Pi))]	
	0.0088	1.008765522		0.993078224	-0.0069
GMALPE =	WMALPE -	[(∑Ej)/(∑Pj)]	+	[∑(Ej/(L*Pj))]	
	0.0088	1.008765522		1.005086696	0.0051
JMALPE =	WMALPE -	[(∑Eij)/(∑Pij)]	+	[∑(Eij/(m*Pij))]	
	0.0088	1.008765522		1.013636434	0.0136

As can be seen in comparison with the values shown in Table 5, the values for RMALPE, GMALPE, and JMALPE derived using equations [7], [8], and [9], respectively, are the same.

III. The Relationship between Dimensional MALPEs and the Joint MALPE

In this section, I first will cover the relationship between the dimensional MALPEs and then the relationship between them, on the on hand, and the joint MALPE, on the other.

III.A The Relationship between Dimensional MALPEs.

Although this can be generalized to more than two dimensions and to more than the number of categories in the preceding example data set using Race (3 categories) and Geography (4 categories), I will stick with this example for now. As you can see in Table 1, the estimates for RACE are summarized by the "marginal total" for each of the 3 columns (i) and the estimates for GEOG are summarized by the "marginal total" for each of the 4 the rows (j).

As defined before, let RMALPE = MALPE for RACE

and GMALPE = MALPE for GEOG

JMALPE = MALPE for RACE & GEOG JOINTLY. i.e., the overall ERROR, which is

∑(Eij – Pij)/(Pij))

WMALPE = weighted MALPE

where WRMALPE = Weighted RACE MALPE

WGMALPE = Weighted GEOG MALPE

WJMALPE = Weighted JOINT (overall both i and j) MALPE

[An important note is that WRMALPE = WGMALPE = WJMALPE. I use this in actually deriving the relationships shown below between RMALPE and GMALPE. The algebra for the derivations is tedious and not shown here. The details are in the attached excel file, along with numerical examples and definitions of RMALPE, GMALPE, and the WMALPEs].

Then

$$RMALPE = [\sum(Ei/k*Pi)] - [\sum(Ej/L*Pj)] + GMALPE \quad [10]$$

$$GMALPE = [\sum(Ej/L*Pj)] - [\sum(Ei/k*Pi)] + RMALPE \quad [11]$$

where Ei = Estimate for Race Group i

k = number of i categories (columns)

Pi = Actual number for Race Group i (Census)

Ej = Estimate for geography group j

L = number of j categories (rows)

We can interpret these relationships as follows:

(1) RMALPE is equal to GMALPE +

(the sum of: (the estimates by racial group, which are in the i column marginals of the estimates table divided by (the product of corresponding population numbers for each racial group, which are in the i column marginals of the population table and the number (k) of column marginals (racial groups)))

-

(the sum of: (the estimates by geography, which are in the j row marginals of the estimates table divided by (the product of the corresponding population numbers for each geographic group in the j row marginals of the population table and the number (L) of row marginals (geographic groups))).

and

(2) GMALPE is equal to RMALPE +

(the sum of: (the estimates by geography, which are in the j row marginals of the estimates table divided by (the product of the corresponding population numbers for each geographic group in the j row marginals of the population table and the number (L) of row marginals (geographic groups)))

-

(the sum of: (the estimates by racial group, which are in the i column marginals of the estimates table divided by (the product of corresponding population numbers for each racial group, which are in the i column marginals of the population table and the number (k) of column marginals (racial groups))).

Using the data in Tables 1 and 2, we find that the empirical data results confirm the statements in equations [10] and [11]. Recall, again, from Table 5 that

Dimension	MALPE
RACE	-0.0069
GEOG	0.0051
JOINT	0.0136

Using equations [10] and [11] on the data in tables 1 and 2, we get the results shown in Table 7.

Table 7. Relationship Between RMALPE and GMALPE

RACE = Dimension i
GEOG = Dimension j
JOINT= Dimension ij

RACE	MALPE =	[∑(Ei/kPi)] -	[∑(Ej/kPj)] +	GMALPE =	RMALPE
		0.993078	1.005087	0.0051	-0.0069
GEOG	MALPE =	[∑(Ej/kPj)] -	[∑(Ei/kPi)] +	RMALPE =	GMALPE
		1.005087	0.993078	-0.0069	0.0051

As you can see in Table 7, the empirical results correspond to what is expected from equations [10] and [11] as given in Table 5.

III.b The Relationship between Dimensional MALPEs and the Joint MALPE.

As you would suspect, a joint MALPE can be expressed in terms of each of the dimensional MALPEs. Continuing with our example of an ex post facto evaluation of error on two dimensions, Race and Geography, we will describe the relationship of RMALPE to JMALPE and the relationship of GMALPE to JMALPE. First, note that we can express JMALPE in terms of the Weighted MALPE (which remember are all equal: the dimensional WMALPEs are equal to one another and to the JMALPE).

JMALPE = WMALPE - [(∑Eij)/(∑Pij)] + [∑(Eij/(m*Pij))] [12]

With the help of equation [12] and some manipulation of terms, I first express JMALPE in terms of RMALPE [13] and then in terms of GMALPE [14]:

JOINT MALPE = [∑(Eij/(m*Pij))] - [∑(Ei/(k*Pi))] + RMALPE [13]

JOINT MALPE = [∑(Eij/(m*Pij))] - [∑(Ej/(L*Pj))] + GMALPE [14]

where, as before

k = number of columns
L = number of rows
M = number of cells (k*L)

Using equations [13] and [14] on the data in tables 1 and 2, we get the results shown in Table 8.

Table 8. The relationship between Dimensional MALPES and their Joint MALPE

JOINT	MALPE =	[∑(Eij/(m*Pij))]	-	[∑(Ej/(L*Pj))]	+	GMALPE	=	JMALPE
		1.013636434		1.005086696		0.0051		0.0136
JOINT	MALPE =	[∑(Eij/(m*Pij))]	-	[∑(Ei/(k*Pi))]	+	RMALPE	=	JMALPE
		1.013636434		0.993078224		-0.0069		0.0136

As you can see in Table 8, the empirical results correspond to what is expected from equations [12], as given in Table 5.

To summarize,

(1) JMALPE is equal to RMALPE +

the sum of: (the estimates by the 12 race-geography groups in the estimates table, divided by (the product of the corresponding population numbers for each of the 12 race-geographic groups of the population table and the number (m=12) of cells comprising the 12 race-geographic groups))). -

(the sum of: (the estimates by racial group, which are in the i column marginals of the estimates table divided by (the product of corresponding population numbers for each racial group, which are in the i column marginals of the population table and the number (k=3) of column marginals (racial groups)))

and

(2) JMALPE is equal to GMALPE +

the sum of: (the estimates by the 12 race-geography groups in the estimates table, divided by (the product of the corresponding population numbers for each of the 12 race-geographic groups of the population table and the number (m=12) of cells comprising the 12 race-geographic groups))). -

(the sum of: (the estimates by geography, which are in the j row marginals of the estimates table divided by (the product of the corresponding population numbers for each geographic group in the j row marginals of the population table and the number (L=4) of row marginals (geographic groups)))

As a way of summarizing these results to-date, here is my answer to someone who asks why the MALPEs in Table 5 are different and, given these differences, how they are related to one another:

"Let me answer this question by looking first at JMALPE, which is .0136. One might think that it is an average of RMALPE (-.0069) and GMALPE (.0051), but it is not. The relationship between, on the one hand, RMALPE and GMALPE, and, on the other, JMALPE is a bit more complicated in that it involves the implicit "weights" stemming from the estimates and populations for the 3 race groups and the 4 geographic areas, which together translate into 12 (4 *3) race-geography groups. We have to take into account these groupings as well as the estimates and actual populations within them.

Once we do this, we can express JMALPE as a function both of RMALPE and of GMALPE, as follows.

$$JMALPE = RMALPE + [\sum(Eij/(12*Pij))] - [\sum(Ei/(3*Pi))]$$

$$0.0136 = -0.0069 + 1.013636434 - 0.993078224$$

$$JMALPE = GMALPE + [\sum(Eij/(12*Pij))] - [\sum(Ej/(4*Pj))]$$

$$0.0136 = 0.0051 + 1.013636434 - 1.005086696$$

Thus, JMALPE is found by adding to RMALPE the sum of: (the estimates by the 12 race-geography groups in the estimates table, divided by (the product of the corresponding population numbers for each of the 12 race- geographic groups of the population table and the number (m=12) of cells comprising the 12 race-geographic groups))) and then subtracting from RMALPE the sum of: (the estimates by racial group, which are in the i column marginals of the estimates table divided by (the product of corresponding population numbers for each racial group, which are in the i column marginals of the population table and the number (k=3) of column marginals (racial groups))).

In terms of GMALPE, JMALPE is found by adding to GMALPE the sum of: (the estimates by the 12 race-geography groups in the estimates table, divided by (the product of the corresponding population numbers for each of the 12 race- geographic groups of the population table and the number (m=12) of cells comprising the 12 race-geographic groups))) and then subtracting from GMALPE the sum of: (the estimates by geography, which are in the j row marginals of the estimates table divided by (the product of the corresponding population numbers for each geographic group in the j row marginals of the population table and the number (L=4) of row marginals (geographic groups))).

RMALPE is related to GMALPE by adding to GMALPE, the sum of: (the estimates by geography, which are in the j row marginals of the estimates table divided by (the product of the corresponding population numbers for each geographic group in the j row marginals of the population table and the number (L) of row marginals (geographic groups))) and subtracting from RMALPE the sum of: (the estimates by racial group, which are in the i column marginals of the estimates table divided by (the product of corresponding population numbers for each racial group, which are in the i column marginals of the population table and the number (k) of column marginals (racial groups))).

Finally, GMALPE is related to RMALPE as follows. To get GMALPE we would add to RMALPE, the sum of: (the estimates by geography, which are in the j row marginals of the estimates table divided by (the product of the corresponding population numbers for each geographic group in the j row marginals of the population table and the number (L) of row marginals (geographic groups))) and then subtract from RMALP, the sum of: (the estimates by racial group, which are in the i column marginals of the estimates table divided by (the product of corresponding population numbers for each racial group, which are in the i column marginals of the population table and the number (k) of column marginals (racial groups))).

IV. Some Concluding Thoughts

It appears that in two dimensions (e.g., race and geography), we can answer the question: "if one was not only doing ex post facto evaluations of estimates (or projections) of the total population but by characteristics such as race and geography, how would one summarize the summary measures of error?" The paper has examined this issue in two dimensions using a standard summary measure of ex post facto error “Mean Algebraic Percent Error” (MALPE). The paper finds that relationships do exist not only between the MALPEs on two dimensions, but between them and their “joint” (overall) MALPE. offers proofs of these relationships both mathematically and empirically. Like most findings, these suggest additional research.

1. The “weighting” of MALPE shares some features with a common demographic technique, “direct standardization.” However, it is not the same. The common features may, however, yield a readily interpretable “decomposition” of the difference between MALPE on one dimension (e.g., Race) and on another (e.g., geography) for the same set of estimates/forecasts (Kitagawa, 1955; Das Gupta, 1978).

2. It is intriguing that WMALPE, which is, in fact, a weighted mean, ends up being the same as the “overall percent error.” In the hypothetical population, the overall percent error is ((E –P)/P) = (1320-1369)/1369) = -49/1369 = -0.0358. This is the same result we get with WMALPE (the weighted mean percent error).

3. What about MAPE and, especially, Root Mean Square Error (RMSE), both other common summary measure of error? Exploring RMSE may prove fruitful because it uses Sum of Squared Errors, a concept ubiquitous in statistics and an important tool in the analysis of “error.”

4. Can the relationships for MALPE be extended to more than two dimensions? The results suggest that this will be the case, but both mathematical proofs and corresponding empirical examples are needed.

Endnotes

1. This paper stems from a question George Hough asked me, which was "if one was not only doing ex post facto evaluations of estimates (or projections) of the total population but by characteristics such as race and geography, how would one summarize the summary measures of error (e.g., MALPE by Race, MALPE by geography and MALPE for the total population)?" This question immediately got my attention and I contacted demographers I knew who were experienced in doing ex post evaluations of estimates and projections. Their responses were similar to mine, not only had not one of them ever thought of it, but also none of them knew of anybody who had dealt with this issue. In addition to the question posed by George Hough, others that have provided helpful

feedback on earlier drafts of this paper include, Jack Baker, Matt Kaneshiro, Stan Smith, Richard Verdugo, Paul Voss, and Paul Wilson,

References

Das Gupta, P. 1978. “A general method of decomposing a difference between two rates into several components.” *Demography* 15(1): 99-112.

Kitagawa, E. 1955. “Components of a difference between two rates.” *Journal of the American Statistical Association* 50 (272): 1168-1194.

APPENDIX A.

PROOF: MALPE = WMALPE + [$\sum$((Ei/(kPi))] - [($\sum$Ei)/($\sum$Pi)]

We know from Equation [1.e] that

$$\text{MALPE} = [\sum((Ei)/(k^*Pi))] - 1 \qquad [A.1]$$

And from Equation [2.d] we know that

$$\text{WMALPE} = [\sum Ei/\sum Pi] - 1 \qquad [A.2]$$

Since we want to know how WMALPE is related to MALPE, let

$$\text{WMALPE} = \text{MALPE} + X \qquad [A.3]$$

Which is equal to

$$[\sum Ei/\sum Pi] - 1 = [\sum((Ei)/(k^*Pi))] - 1 + X \qquad [A.3.1]$$

Rearranging terms to isolate X, we get

$$[\sum Ei/\sum Pi] - 1 - [\sum((Ei)/(k^*Pi))] + 1 = X \qquad [A.3.2]$$

Which reduces to

$$[\sum Ei/\sum Pi] - [\sum((Ei)/(k^*Pi))] = X \qquad [A.3.3]$$

And substituting X into A.3 we get

$$\text{WMALPE} = \text{MALPE} + [\sum Ei/\sum Pi] - [\sum((Ei)/(k^*Pi))] \qquad [A.4]$$

Which means that

$$\text{MALPE} = \text{WMALPE} + [\sum((Ei)/(k^*Pi))] - [\sum Ei/\sum Pi] \qquad [A.5]$$

And we have our proof that

$$\text{MALPE} = \text{WMALPE} + [\sum((Ei/(k^*Pi))] - [(\sum Ei)/(\sum Pi)]$$

APPENDIX B.

PROOF: $JMALPE = WJMALPE - [(\sum Eij)/(\sum Pij)] + [\sum(Eij/(m*Pij))]$

This is for only two dimensions, i and j, with i=1 through k and j=1 through L

Since we know from equation [1.b] that on any given single dimension (e.g., i=1 through k) that

$$MALPE = [1/(\sum Pi/k)/\sum Pi]*[\sum (Ei-Pi)/Pi] \qquad [B.1]$$

And from Appendix A, we know that

$$WMALPE = MALPE + [(\sum Ei)/(\sum Pi)] - [\sum(Ei/kPi)] \qquad [B.2]$$

Since

$$MALPE = [\sum((Ei)/(k*Pi))] -1 \qquad [B.3]$$

and

$$WMALPE = [\sum Ei/\sum Pi] – 1 \qquad [B.4]$$

We can generalize [B.1] to two dimensions

$$MALPE = [1/(\sum Pij/k*L)/\sum Pij]*[\sum(Eij-Pij)/Pij] \qquad [B.5]$$

Let JMALPE stand for a MALPE on 2 dimensions, i and j and m = k*L, so we can re-express [B.5] as

$$JMALPE = [1/(\sum Pij/m)/\sum Pij]*[\sum(Eij-Pij)/Pij] \qquad [B.6]$$

Using Equations [1.a] through [1.e] in conjunction with [B.6], we have

$$\text{Since } (1/m) = [1/(\sum Pij/m)/\sum Pij] \qquad [B.6.1]$$

$$JMALPE = [1/(\sum Pij/m)/\sum Pij]* [\sum (Eij-Pij)/Pij] \qquad [B.6.2]$$

$$= /(\sum(Pij/m)/\sum Pij]*[\sum (Eij-Pij)/Pij] \qquad [B.6.3]$$

$= [\sum((Eij-Pij)/(m*Pij))]$ [B.6.4]

$= [\sum((Eij)/(m*Pij))] -1$ [B.6.5]

$WJMALPE = \sum [(Pij/\sum Pij)*((Eij-Pij)/Pij)]$ [B.7]

NOTE: WJMALPE also equals WMALPE on any dimension (i or j)

And since

$WJMALPE = \sum [(Pij/\sum Pij)*((Eij-Pij)/Pij)]$ [B.7.1]

$= \sum[Eij-Pij)/\sum Pij]$ [B.7.2]

$= \sum Eij/\sum Pij - \sum Pij/\sum Pij$ [B.7.3]

$= [\sum Eij/\sum Pij] – 1$ [B.7.4]

Thus, we have

$JMALPE = [\sum((Eij)/(m*Pij))] -1$ [B.8]

and $1 + JMALPE = [\sum((Eij)/(m*Pij))]$ [B.8.1]

$WJMALPE = [(\sum Eij)/(\sum Pij)] -1$ [B.9]

and $1 + JWMALPE = [(\sum Eij)/(\sum Pji)]$ [B.9.1]

Since we want to know how WJMALPE is related to JMALPE, let

$WJMALPE = JMALPE + X$ [B.10]

Which is equal to

$[\sum Eij/\sum Pij] – 1 = [\sum((Eij)/(m*Pij))] -1 \quad + X$ [B.10.1]

Rearranging terms to isolate X, we get

$[\sum Eij/\sum Pij] – 1 - [\sum((Eij)/(m*Pij))] + 1 \quad = X$ [B.10.2]

Which reduces to

$[\sum Eij/\sum Pij] - [\sum((Eij)/(m*Pij))] = X$ [B.10.3]

And substituting X into [B.10.3] we get

$WJMALPE = JMALPE + [\sum Eij/\sum Pij] - [\sum((Eij)/(m*Pij))]$ [B.11]

Which means that

$JMALPE = WJMALPE + [\sum((Eij)/(m*Pij))] - [\sum Eij/\sum Pij]$ [B.11.1]

And we have our proof that

$JMALPE = WJMALPE + [\sum(Eij/(m*Pij))] - [(\sum Eij)/(\sum Pij)]$