

1 To: TLUMIP gang, Bill Upton, cmc3@ns.algarnet.net (Doug Hunt in Brazil)
2 From: Rick Donnelly <Rick.Donnelly@worldnet.att.net>
3 Subject: Utility scaling-Resp from T de la Barra
4 Date: Tue, 22 Oct 96 22:34:18 +0000
5 X-Attachments:

6
7 The attached reply to our collective comments about the (dis)utility
8 scaling issue came this evening from Tomas de la Barra. I had sent him a
9 compilation of the recent email traffic on this topic last Friday evening
10 (18 October), asking him to comment as he saw fit. I still want to see
11 his revised paper, which he should have on his Web site by tomorrow even-
12 ing (Wednesday, 23 October) and to review these comments at length. But
13 if he implements the scaling parameter he mentions below, then I'd say
14 that this becomes a moot issue. I'd appreciate receiving comments from
15 any of you on this topic, but would like to request them (or at least a
16 notice that you're preparing them, and a brief synopsis of them) by this
17 Friday (25 October).

18
19 Happy reading!

20
21 >From: Modelistica T del Barra <73000.1534@CompuServe.COM>
22 >To: Rick Donnelly <Rick.Donnelly@worldnet.att.net>
23 >Subject: Re: Oregon study
24 >Date: Tue, 22 Oct 96 20:30:47 +0000

25 >
26 >Dear Rick,
27 >
28 >Thanks for your message and for sending me the set of comments from
29 >members of the peer review panel. They are, no doubt, very interesting
30 >comments from experienced and knowledgeable people. These ideas will help us
31 >to improve and strengthen our postulates, so that I'll be happy to reply
32 >to every one of them. I trust that the committee is not concentrating
33 >only on the scaled utilities issue, and that they are giving due
34 >consideration to all other aspects of modeling, such as the ones I
35 >included in my original notes to Paul. Scaled or not, it is important
36 >that the model is doing a good job at calculating utilities themselves,
37 >and that it is able to provide the analyst with sound tools to represent
38 >travel behavior, intermodality, multipath capabilities, comprehensive
39 >economic calculations, and the like. I also made a lot of emphasis on the
40 >software environment, because after all these years in practice I know
41 >how important it is to the practitioner to have all facilities at the
42 >touch of a button (or mouse click). These are, perhaps, trivial things,
43 >but they allow you to do a better job in less time. You probably have
44 >good experience on this, but it can be very frustrating when you have to
45 >struggle with the programs instead of struggling with the problems. If
46 >people are interested, I can send you a demo version of the Sacramento
47 >application to play around with, like the one on the Web but with real
48 >data and lots of scenarios. Perhaps you can ask for something similar
49 >with the Meplan application to Sacramento. I can think of no better way
50 >of evaluating a model than to have a first hand experience like this one.

51 >
52 >With respect to your question "whether you can (and would be willing) to
53 >provide us with a version of TRANUS that did not employ (dis)utility
54 >scaling if we come to that?" We can do even better than this. You might
55 >have noticed that in the 'conclusions and further work' section of my
56 >paper on improved logit formulations, it is mentioned that the ideal
57 >model would be one in which the degree of scaling could be parameterized.
58 >This can be achieved simply by raising the minimum cost denominator to
59 >the power of a parameter 'a' with a range from zero to one. The utility
60 >term would thus become:

61 >
62 >
$$U = V_i / (\min V_j)^{**a},$$

63 >

64 >where 'Vi' is the cost of an option and 'min Vj' is the cost of the best
65 >option. If you set the value of a=1, you get a fully scaled utility
66 >function; if you set a=0 the model is fully unscaled and identical to the
67 >traditional logit. This model form will allow you to experiment with

68 >different values of 'a' , that is, with different degrees of scaling, or
69 >no scaling at all. In this way you get a more general model, because it
70 >includes the standard logit as a special case.
71 >
72 >This should only take a couple of hours of programming and is very
73 >straightforward to test: just put a value of a=1, rerun and make sure the
74 >model produces exactly the same results as before. Then make runs with
75 >a=0 and some intermediate values, and confirm that the results are
76 >sensible.
77 >
78 >This suggestion in no way means pulling back on the issue of scaling. As
79 >I mentioned, this is suggested in my paper (the little formula up there
80 >is in a footnote in the conclusions), and I think this is the most
81 >sensible thing to do. I agree with the comment by one of the reviewers
82 >that in some cases, like freight movements, scaling might not be
83 >justified (or less justified), if you interpret scaling as a
84 >psychological thing more to do with the behavior of people than firms.
85 >Also, if you want to do a systematic comparison between scaling and
86 >unscaling and intermediates, but making sure that everything else is kept
87 >exactly the same, you need to have this flexibility. In fact we are
88 >implementing it to be ready for applications in Bogota and Swindon (UK)
89 >that are now entering the calibration stage.
90 >
91 >Please circulate the above, because I feel that in this way most concerns
92 >will be satisfied, I hope. Anyway, I attach a bunch of notes on the comments
93 >made by the members of the panel, following ///. The revised paper will
94 >be up in the Web by tomorrow night.
95 >
96 >Best regards,
97 >
98 >Tomas
99 >
100 >PD: Please do not compare scaling to k factors!
101 >
102 >+++++
103 >+-----+
104 >| Comments from Carl Batten (consultant) |
105 >+-----+
106 >Date: Fri, 11 Oct 96 22:52:05 +0000
107 >
108 >If I understood Doug correctly, the "scaling" he is most concerned with
109 >is different than applying a scalar (i.e., constant) to the disutilities
110 >in the logit function. Tranus lets you do that--it has a dispersion
111 >parameter (the betas in Equation 12 on page 11 of the Reference Manual)
112 >for each sector that scales each zone-to-zone disutility measure by a
113 >fixed amount for that sector.
114 >
115 >/// Yes, this is correct.
116 >
117 >But what's really wierd in Tranus is that each individual disutility
118 >measure is divided by the minimum (best case) disutility measure for all
119 >the production zones that could supply a consumption zone, for that
120 >sector. (See Equation 11 on page 11 of the Reference Manual.) So the
121 >"scalar" is different for each consumption zone and can be different
122 >between iterations for the same consumption zone if prices change between
123 >iterations, as TRANUS allows.
124 >
125 >/// Yes, this is also true, although this effect will be slight,
126 >particularly in the last iterations. The scaling will also change if you
127 >introduce any changes to the choice set, such as a new highway, or new
128 >land, taxes, etc.
129 >
130 >The net effect, as nearly as I can understand it, is that disutilities
131 >are no longer expressed in any consistent units such as dollars or
132 >"utils," but are measured in units unique to the consumption of each
133 >commodity in each zone, where the minimum disutility for that commodity
134 >in that zone is always scaled to exactly one and all others expressed as

135 >a multiple of the minimum.
136 >
137 >/// Remeber scaling is only used when calculating choice probabilities.
138 >Also please note that, after scaling, composite costs are un-scaled back
139 >after the log-sum calculation is made, so you get back to dollars or
140 >whatever 'consistent' units you were working with. In fact, utilities
141 >become independent from the units in which you measure utility. This has
142 >the additional advantage that you can 'borrow' parameter values from one
143 >application to another, to use them as starting points.
144 >
145 >The concern Doug expressed about applying a constant dispersion parameter
146 >to disutilities that have been scaled by varying amounts seems quite
147 >valid.
148 >
149 >/// Scaling provides much more flexibility to the model, particularly if
150 >you look at the form with variable scaling in my message to Rick. We have
151 >found that this works particularly well in large-scale modeling, such as
152 >state-wide. It is in such cases that differences in scale are most
153 >noticeable, because you have to deal simultaneously with very short and
154 >very long travel options within the same model. Look at this as a way to
155 >avoid segmentation of the model, which is what people are forced to do in
156 >practice; in the modal split example, segmentation means that you would
157 >have to calibrate several parameters for short, intermediate and long
158 >trips. With scaling you don't have to do this. This is even more
159 >important for location models, where segmentation is not possible.
160 >Consider the example in my paper on the matter ("Improved logit
161 >formulations...") in which we are simulating a residential location
162 >choice, and say options consisted on zones at costs of 3, 8, 20 and 25
163 >from the place of work. People would consider option 3 as very close,
164 >option 8 as intermediate, and both options 20 and 25 as almost equally
165 >distant. The argument here is that people would perceive the difference
166 >8-3=5 as much bigger than the difference between 25-20=5, i.e. perception
167 >is getting smaller per unit of cost as you go further away. Imagine the
168 >set had many more options, including some costing 140 and 145. The
169 >standard un-scaled logit does not make this distinction.
170 >
171 >I doubt that scaling by the minimum is done to prevent the logsums from
172 >going negative. If an individual disutility were negative (which could
173 >happen only if price or distance were negative), that would really mess
174 >things up in TRANUS. The negative disutility would be the minimum, so its
175 >scaled value would be one ($-U/-U = 1$), but the others' scaled values
176 >>would be negative (something positive divided by $-U$), making them each
177 >appear to be better options (lower disutility value) than the best.
178 >
179 >/// You are absolutely right here. The scaling does not prevent the
180 >log-sum from going negative, and this is a separate problem. You are also
181 >right in that if you let negatives into the formulation, it would mess
182 >things up in the next scaled logit up the chain. But negative log-sums
183 >are wrong anyway, as you correctly suggest, and our composite cost
184 >formulation avoids this altogether.
185 >
186 >+-----+
187 >| Comments from Carl Batten (consultant) |
188 >+-----+
189 >
190 >I also downloaded de la Barra's paper he presented in Sweden on why his
191 >scaling is a good thing and am trying to work through it. He contends
192 >that people don't behave as if they evaluate alternatives in terms of
193 >absolute differences in utilities, but rather differences in relative
194 >utilies and that his method therefore has better predictive power and is
195 >easier to calibrate. He claims to prove that his model is indeed a logit
196 >and satisfies IIA.
197 >
198 >/// Now there is an improved version of this paper up in the Web. The
199 >argument on why the model continues to be a logit and satisfies IIA has
200 >been refined and completed. The main point here is that, if we want to
201 >scale utility, any arbitrary positive constant would do the job. We

202 >picked up the cost of the best option, but we could have used any other,
203 >such as the mean cost. For any given choice set, if you divide utilities
204 >by a constant term, it is easy to prove that you can maintain
205 >homoscedasticity (equal variance) in the Gumbel distribution of the error
206 >terms, hence you end up with a logit with IIA. What is not guaranteed is
207 >that the variance is kept across different choice sets, and I haven't
208 >seen anyone defending this in the literature. It is argued from the
209 >example of Figure 2 in the paper that mode choice for destination 1 is
210 >bound to have a different variance than the mode choice to destination 2.
211 >In fact, Hugh Williams showed that the value of the logit parameter
212 >should be approx. equal to the mean cost, so that if 0.2 is a reasonable
213 >parameter for destination 1, then 0.002 should be used for destination 2
214 >for consistency. This is exactly what we are doing; dividing the utility
215 >by the mean cost (or the minimum) is equivalent to dividing the parameter
216 >itself. Note that IIA will also change if the scale of the choice is
217 >changed. If you have a set with options costing 20 and 25, people would
218 >evaluate these options and think that 20 is much better than 25. But if
219 >you then introduce a new option costing only 5, then the whole scale of
220 >the choice is changed, and people will consider that 20 and 25 are
221 >equally bad. IIA is meant to be a property for a single specific choice,
222 >but no one can claim that the same variance should be maintained
223 >throughout all choices. Precisely for this reason we have hierarchical
224 >and chained conditional logits: we want to be able to use different
225 >dispersions in each case. Scaling is just a practical way to simplify the
226 >modeling process, avoiding segmentation.
227 >
228 >/// The following comment is, I think, from Doug to Carl
229 >
230 >I have received something you wrote regarding the scaled logit function
231 >in TRANUS in response to an email from Rick. I haven't seen the email by
232 >Rick or the comments by Tim or Pat to which you refer. In any case, you
233 >are on the right track: I am concerned about where the disutility measure
234 >for allocation of production to each zone is divided by the minimum (best
235 >case) disutility measure for all the production zones, as in Equation 11
236 >on page 11 of the Reference Manual. Your description of the effect and
237 >the resulting concerns is correct.
238 >
239 >/// I haven't received anything from Tim or Pat either.
240 >
241 >You also consider the situation where the disutilities go negative. You
242 >are correct, except where you indicate that this could only happen when
243 >the costs or times are negative. Actually, there is a constant value
244 >included in the utility function that can be highly negative or positive.
245 >A logsum value for a set of positive disutilities can also 'go negative'
246 >if the dispersion parameter is small enough. It seems to me that a switch
247 >in sign is readily possible, which the standard logit 'takes in stride'.
248 >I honestly don't know how Tomas deals with this. Maybe he has special
249 >flags that warn when there are changes in sign...
250 >
251 >/// No flags, thank you. The log-sum tends to go negative as the value of
252 >the parameter gets smaller, and quite rapidly, in fact. This is also
253 >shown in a plot diagram in my paper. Many people have written against
254 >this property of the log-sum, most notably Fisk and Boyce (referenced in
255 >my paper). Their argument is that the negativity property is inconsistent
256 >with the interpretation of the log-sum as an average cost. We have added
257 >to this the argument that the logit model is being used as a demand
258 >model, so that prices and quantities consumed only exist in the positive
259 >quadrant of the demand function. The standard argument in favor of
260 >letting the log-sum to go negative is that it doesn't matter, because it
261 >is only a utility indicator, and that it is only the difference that
262 >counts when comparing two situations. This, however, is a problem if the
263 >log sum is used as the average cost further up the hierarchy.
264 >Furthermore, as I show in the revised version of my paper, the log-sum
265 >tends to overestimate benefits. Anyway, we use a different form of
266 >calculating composite costs, and we have shown that the results are
267 >always positive or zero, and that benefits are not being overestimated.
268 >

```
269 >+-----+
270 >| Comments from Carl Batten (consultant) |
271 >+-----+
272 >
273 >Randy and I reviewed de la Barra's paper about his "improved logit
274 >formulation." We concluded that it has some serious problems from the
275 >standpoint of economic theory. The foundation for his scaling is his
276 >interpretation of the fundamental microeconomic principle of diminishing
277 >marginal utility whereby the more you consume, the less you enjoy each
278 >additional unit of consumption. He claims that the usual way of doing a
279 >logit model is not consistent with this theory because it treats the
120 >value of an additional dollar or minute as being the same whether it is
281 >in the context of a short, cheap trip or a long, expensive one. In the
282 >case of goods shipments he is certainly wrong about that and in the case
283 >of people movements he is generally wrong. To a producer buying
284 >intermediate goods, a dollar spent on transport IS worth the same as any
285 >other dollar, whether it is part of the cost of a million dollar order or
286 >part of the cost of a ten dollar order. To a consumer making a trip, an
287 >extra dollar spent on a cheap trip is worth just about the same as a
288 >dollar saved on an expensive trip. Wanting or needing to take an
289 >expensive trip does not raise one's income, so total consumption does not
290 >increase, so the marginal utility of the marginal dollar does not change
291 >as one must substitute out of the consumption of some other good, the
292 >marginal value of a dollar's worth of which was, at the previous instant,
293 >worth the same as the marginal dollar's worth of travel. De la Barra
294 >seems to confuse the marginal utility of an additional unit of travel
295 >with the marginal utility of an additional dollar.
296 >
297 >What de la Barra does, though, goes way beyond addressing diminishing
298 >marginal utility. He assumes that the marginal value of a dollar (or
299 >minute) is directly proportional to the cost of each trip. So a dollar
300 >spent on a $10 trip is worth 10 times as much as a dollar spent on a $100
301 >trip. This is doesn't fit with either consumer theory or my observation
302 >of the real world. While the impact of unobserved variables is more
303 >likely to swamp the impact of one dollar's worth of an included variable
304 >when modeling a $100 trip than when modeling a $10 trip, a dollar is
305 >still worth a dollar to the consumer. The consumer can buy the same
306 >number of ice cream cones with a dollar he saved on a $10 trip as with a
307 >dollar he saved on a $100 trip.
308 >
309 >/// The argument is 'marginally decreasing perception of utility', which
310 >is quite different to the Marshallian law of diminishing marginal utility
311 >(nowadays, superseded by Edgeworth's Indifference Analysis), but doesn't
312 >apply to travel and location anyway, because transport is a derived
313 >demand. What you can consume at the end of the trip is what really
314 >matters, and this is not included in the utility function. In this sense,
315 >you might not want to trade off an extra $1 ice cream cone for an
316 >additional $1 trip to a shop, but if it was related to a $500 trip to the
317 >Bahamas for vacation, the dollar difference will tend to vanish. I agree,
318 >though, that when it comes to production, transport is inserted more in
319 >the world of real economics, than in psychological perceptions. This is
320 >the reason why I think that a logit with variable scaling is more
321 >flexible. If you are simulating the flow of goods from, say, agriculture
322 >to industry, you may want to use very little scaling or none at all. If
323 >you are simulating work-home relationships, you might get better results
324 >with scaling. The problem with economics (and what makes it so
325 >interesting) is that there is a lot of fussy arguments and it is
326 >sometimes difficult to agree on certain aspects (we could argue endlessly
327 >on the concept of utility, for instance). This is why I try to avoid
328 >sentences such as Mr A is right, and Ms B is completely wrong.
329 >
330 >The trick de la Barra uses to claim that his formulation is still a logit
331 >is unlikely to hold in any real situation. To remain a logit after his
332 >scaling, the errors must have already been distributed in such a way that
333 >the ratio of unexplained to explained variation is exactly proportional
334 >to the disutility level across short and long trips. How likely is that?
335 >
```

336 >/// I am not trying to trick anyone. I have already described the logic
337 >behind this proof. If you adopt homoscedastic Gumbel distributions of the
338 >error terms, you get a logit model after integration. The fact that you
339 >divide utility by a constant to set the scale does not change this. What
340 >is quite obvious is that if you change the scale, variability will also
341 >change in this formulation. This, I think, is a more flexible model, and
342 >a very practical solution to avoid complex segmentations if the scales
343 >are really affecting your results. The model with variable scaling has
344 >the advantage of including the standard logit as a special case. Imagine
345 >you are calibrating a modal split model, using max likelihood. Pick a
346 >sub-sample of short distance trip-makers and another sub-sample of long
347 >distance trip-makers. You are bound to get different parameter values,
348 >and you end up segmenting the model. If you do get similar parameters,
349 >then segmentation was not necessary. Because we got a bit tired of doing
350 >this, we thought that scaling was a very practical way of doing this, as
351 >it turned out. If you are doing, for instance, an inner-area study to
352 >improve bus services, then probably scaling is of little relevance and
353 >you might do away with it. If, however, you are doing a whole
354 >metropolitan region or a state-wide model, then you will probably give it
355 >a try to avoid segmentation.
356 >
357 >+-----+
358 >| Comments from Doug Hunt (peer review panelist) |
359 >+-----+
360 >
361 >On Wed, 16 Oct 1996 09:21:37 -0700 lau
362 >
363 >1: Tomas indicated the following regarding the scaled logit formulation:
364 >> It is easy to see that the scaled model is much easier to calibrate
365 >> than the traditional formulation and the results are bound to be better
366 >> and more reliable by a long stretch.
367 >
368 >These are very strong claims that get to the heart of the matter. Is
369 >there any empirical evidence to support them, such as some controlled
370 >comparison tests?
371 >
372 >/// I elaborated this argument a bit better in the revised paper. What
373 >people do when faced with large-scale models is to segment the model
374 >into, say, short, medium and long distance commuters, and make separate
375 >calibrations. This is what CIS was doing in Chile to simulate tolls.
376 >Observing before-and-after data around a tolled highway, including
377 >stated preference surveys, they realized that people doing long trips
378 >were prepared to pay more. O-D surveys indicated that after the new
379 >tolled highway, users of the toll showed longer mean trip lengths. People
380 >living and working close to the toll highway were the main users of the
381 >un-tolled alternative. This is why the guys at CIS were using separate
382 >calibrations. When they got our scaled model, they realized that they
383 >could get even better results without having to segment the model.
384 >Calibrating one parameter is, of course, easier than calibrating several,
385 >and reduces the size of the sample. Have we made controlled and rigorous
386 >comparisons? No, and I mention this in my paper in the 'further work'
387 >section of my paper. Has anybody done similar experiments to show that
388 >scale is irrelevant in most circumstances? The reply to this is, again,
389 >no. What we need here is some clever guy at a university to do this
390 >research, which would make a very interesting project. He or she would
391 >need a model that could vary the degree of scaling.
392 >
393 >2: Tomas also indicated:
394 >> Furthermore, if you are using a maximum likelihood method to estimate
395 >> the parameters, based on stated preference data, standard software such
396 >> as Alogit or Hyellow may be used. It is simply a matter of entering the
397 >> scaled values to the program. This is currently being made by a Belgium
398 >> group to calibrate the Transus housing choice and modal/split assignment
399 >> models.
400 >
401 >It is unclear how the scaled values can be entered to the program. The
402 >scaling factor is a utility value that is calculated using the utility

403 >function, which has parameter values estimated by the program. So one
404 >needs the answer to determine the answer. Are the utility values
405 >calculated beforehand, perhaps by another program? Is the program used
406 >twice: once to get the utility function parameters for a standard logit
407 >model and then again with scaled utility values to get a second
408 >dispersion parameter? In either case the process is not full-information
409 >maximum likelihood. We need more explanation than we have been given so
410 >far.
411 >
412 >/// The Belgians tell me that HieLoW (correct spelling, sorry) is much
413 >more flexible than Alogit for data input. Also, because it's Windows, it
414 >is much easier to interface with Excel or Access, and Transus can produce
415 >final and intermediate outputs in compatible form. This solves the data
416 >problem and you can enter scaled values, but in no way solves the many
417 >other complexities around calibrating with max likelihood. To work around
418 >the 'one needs the answer to determine the answer' problem, you have to
419 >proceed from the bottom up in the chain. This means that you estimate the
420 >logit assignment/modal split first, and then proceed upwards from the
421 >results of the previous stage. When you get to the top, you assign,
422 >capacity-restrain and re-adjust the parameters. This has little to do
423 >with scaling, and any logit-based integrated land use transport model
424 >will have to something like this.
425 >
426 >The situation regarding the alternative specific constants is also
427 >unclear. With the standard logit formulation when there are A
428 >alternatives then the full specification of alternative specific
429 >constants includes A-1 constants. I think that because the scaled logit
430 >is not translationally invariant then the full specification includes A
431 >rather than A-1 constants (although I must admit this is based more on
432 >some intuitive musings while riding in the Brazilian countryside than on
433 >'hard pencil and paper work' so far). I am not sure how a standard logit
434 >estimation package would 'react'.
435 >
436 >/// We usually set the alternative specific constant of the preferred
437 >option to 1, so that you have to estimate A-1 constants. I don't see why
438 >scaling should change this.
439 >
440 >3: Tomas also indicated that the IIA property is preserved with the
441 >scaled logit formulation - and that a proof of this is provided in a
442 >paper.
443 >
444 >I want to see this proof. At this point I cannot agree - and I think the
445 >IIA property is not preserved in general. It is my understanding that the
446 >IIA property is that the ratio between the choice probabilities for two
447 >alternatives is not affected by the presence or absence of a third
448 >alternative. But with the scaled logit, if the alternative with the
449 >maximum utility in one choice set is excluded from another choice set,
450 >then the scaling parameter is different and therefore all the exponents
451 >are different, which means that all the probability ratios are also
452 >different. Consider the following:
453 >
454 >In standard logit:
455 >
456 > $P_a/P_b = \exp(xU_a)/\exp(xU_b)$
457 >
458 >where: P_a = probability that alternative a selected
459 > P_b = probability that alternative b selected
460 > x = dispersion parameter (I can't get a lambda, you can whiteout
461 > the upper right stem if you like ...)
462 > U_a = utility of alternative a
463 > U_b = utility of alternative b
464 >
465 >This demonstrates the IIA property in that the presence or absence of an
466 >alternative c with utility U_c does not affect the ratio P_a/P_b .
467 >
468 >However, in scaled logit:
469 >

```

470 >Pa/Pb = exp(x[Ua/Umax])/exp(x[Ub/Umax])
471 >
472 >                (1/Umax)
473 >    = {exp(xUa)/exp(xUb)}
474 >
475 >where: Umax = the utility of the alternative with the maximum utility out
476 >            of the set of alternatives
477 >
478 >This shows that the utility of a third alternative in general, Umax,
479 >affects the ratio Pa/Pb. If the alternative associated with Umax is
480 >removed from the choice set, then Umax changes and the ratio Pa/Pb
481 >changes. This violates the IIA property as I recall it.
482 >
483 >Perhaps I am missing something. I wish I could see the proof, but I do
484 >not have net reading software available at the moment. And I am working
485 >entirely from my memory (which is far from perfect) on all the
486 >definitions concerned. I suppose the IIA property is maintained with the
487 >scaled logit formulation in the more limited case where the maximum
488 >utility alternative is not changed or is Ua or Ub - which may be the
489 >proof presented in the paper - but that is fixing the defining
490 >characteristic of the scaled logit formulation and therefore is rather
491 >restrictive.
492 >
493 >All this on the IIA property may seem rather academic, but I think that
494 >if the IIA property is not preserved then it means that the definition of
495 >choice sets considered in the model becomes much more critical, which is
496 >a very practical issue indeed.
497 >
498 >/// The proof has been expanded in the revised version of the paper, and
499 >I have already described it above. It is based on the interpretation of
500 >Umax as an arbitrary constant to set the scale. If you divide the utility
501 >values by a constant 'q' in a logit model, then the ratio of one choice
502 >over another is:
503 >
504 >    exp x(Ua/q) / exp x(Ub/q)
505 >
506 >thus retaining the IIA property. This is because in the derivation, the
507 >homoscedastic Gumbel distributions of the error terms have been
508 >maintained, WITHIN A SPECIFIC CHOICE SET. IIA is not preserved, however,
509 >across different choice sets if the scale changes, and this is meant to
510 >solve, or at least alleviate, an important weakness in the standard logit
511 >model. Most authors claim that IIA is a bad property, even within a
512 >choice set. The intention of the scaled model is to alleviate IIA across
513 >choice sets, but unfortunately it doesn't provide a solution to the
514 >within choice problem. In fact, the scaled model will fall into the red
515 >bus-blue bus trap just as easy as the standard logit. We have, however,
516 >solved the attribute correlation problem where it is most acute: in
517 >assignment/modal split.
518 >
519 >+-----+
520 >| Comments from Maren Outwater (consultant) |
521 >+-----+
522 >
523 >Date: Thu, 17 Oct 96 18:55:09 +0000
524 >
525 >...Also, I wanted to throw in my 2 cents worth on the logit scaling
526 >issue. I realize that I am a conservative on this topic...but I concur
527 >with others that the scaling factor in the logit equation violates the
528 >IIA property and that this fact makes the formulation something other
529 >than logit. Theoretically, this makes me nervous because I feel that it
530 >will be difficult to calibrate. In practice, the logit equation works
531 >well because it is not dependent on zone size, but the inclusion of a
532 >scaling factor will change this. Other aspects of traditional models that
533 >are dependent on zone size (or some factor of size) are difficult to
534 >calibrate. The Michigan example comes to mind, where Pat spent mucho
535 >hours trying to deal with the zones outside the state in trip
536 >distribution. Obviously, this is not a logit model, but it raises the

```

537 >question of difficulties in incorporating a scaling factor based on size
538 >in model formulations. I have long been wary of traditional models that
539 >have to deal with these scaling factors on size; fortunately the trend is
540 >towards models that do not have to deal with these scaling factors (ie
541 >logit formulations). If de la Barra has had success using his formulation
542 >in calibrating and applying models, I would be interested to read about
543 >these examples.
544 >
545 >/// This, I think, is a very reasonable 2 cents comment. I disagree,
546 >however, with the argument that 'In practice, the logit equation works
547 >well because it is not dependent on zone size'. In my paper, I included a
548 >section on aggregation to deal precisely with this problem. The error
549 >term in the logit model must explain all sources of variability in the
550 >perception of utility, not included in the utility term itself.
551 >Variations in tastes, imperfect information, poor model specification and
552 >many other non-modelled differences are all absorbed by the distribution
553 >of the error terms. If you have, say, a modal split model, and you split
554 >the travel-making population into two groups, the degree of variability
555 >will be much larger than if you split the population into ten groups.
556 >Because variability is directly linked to the dispersion parameter of the
557 >logit model, if you calibrate the two-groups model you will get a lower
558 >value for the dispersion parameter than with the ten-groups model,
559 >because the former will have to put much more variability into the
560 >distribution of the error terms. This is why in a hierarchical logit you
561 >get $x_1 < x_2 < x_3 \dots$. Zone size is just another source of variability, so that
562 >your parameters will be different in a 50 zones model compared to a 200
563 >zones model.