Shannon's measure of information, path averages and the origins of random utility models in transport itinerary or mode choice analysis

Marc Gaudry, Emile Quinet

▶ **To cite this version:**

HAL Id: halshs-00713168

https://halshs.archives-ouvertes.fr/halshs-00713168

Submitted on 29 Jun 2012

PARIS SCHOOL OF ECONOMICS
ECOLE D'ECONOMIE DE PARIS

Shannon's measure of information, path averages and the origins of random utility models in transport itinerary or mode choice analysis

Marc Gaudry
Emile Quinet

Shannon's measure of information, path averages and the origins of random utility models in transport itinerary or mode choice analysis

Marc Gaudry[1] and Émile Quinet[2]

1   Agora Jules Dupuit (AJD), Département de sciences économiques
    Université de Montréal
    Montréal, marc.gaudry@umontreal.ca

2   Paris-Jourdan Sciences Économiques (PSE)
    École des Ponts ParisTech (ENPC)
    Paris, quinet@enpc.fr

Abstract

We interpret the often mentioned difference between *Logsum* and average utility in terms of Shannon's (1948) information measure S, leading to a **P**ath **A**ggregation **TH**eorem (PATH). It states that, in transport networks where unique measures of the utility of multiple paths are required for demand model formulation purposes and the true path choice model is Multinomial Logit (MNL), constructs based on weighted averages of path characteristics derived from multipath assignments always underestimate the utility of multiple paths, a deficit exactly equal to S (corresponding to minus-one times entropy) if the weights are the path choice probabilities.

We study the properties of this S measure of aggregation error, along with those arising from other types of averages of path characteristics, outlining some implications for demand estimation and project appraisal. Notably, the validity of the PATH does not depend on the specific contents of the representative utility functions (RUF) associated to paths, such as their mathematical form or their eventual inclusion of alternative-generic constants (AGC). We show by simulation that averaging modes or sub-modes — a frequent feature of traffic modeling studies — can lead to important error in terms of level of traffic and welfare measurement.

Concerning the mathematical form of the RUF, we recall that, after the publication of Abraham's 1961 random utility model (RUM) of road path choice deriving the Probit specification based on the Gaussian error distribution (and another specification based on the Rectangular error distribution), French engineers used this seminal approach as justification of road path choice formulae then in current use and assigned the name "Abraham's Law" to a particular standard one, effectively a "Logarithmic Logit" close to the logarithmic RUF carefully specified for Logit mode choice by Warner in 1962. For transit problems, the preference went to a linear RUF, as evidenced in Barbier's casual binomial Probit application to bus and metro, published in 1966, which may have inspired the later generalizations by Domencich and McFadden.

In view of many founders' conscientiously crafted nonlinear Logit formulations, and more generally of the repeatedly demonstrated presence of nonlinearity in RUF path and mode specifications since their careful work 50 years ago, we analyze the impact of such nonlinearity on S. This impact is tractable through a comparison of measures S_2 and S_1 associated with two path choice models differing only in RUF form, as determined by Box-Cox transformations applied to their level-of-service (LOS) variables. We show that, although the difference between measures S_2 and S_1 may reach a minimum or a maximum with changes in LOS, the solution for such a turning point cannot be established analytically but requires numerical methods: the demonstrable impact on S of nonlinearity, or asymmetry of Logit curve response, is tractable, but only at non trivial computational cost.

We point out that the path aggregation issue, whereby aggregation of paths by *Logsums* differs from aggregation of their characteristics by averages, is not limited to public transit (PT) projects with more or less "common" lines competing in dense urban transit networks (our particular Paris predicament motivating the analysis) but also arises in other modes whenever distinct itineraries or lines compete within a single mode. Concerning dense urban PT networks, we hypothesize that *Logsums* based on multiple path assignments treating all transit means (about 10 in our problem) as one modal network should, using Ockham's razor, be simpler than the insertion of a layer of choice hierarchies among such urban means based on non nested specifications embodying assumptions on the identity of "higher" and "lower" means, the latter reasserting the multiple path access problems the hierarchies were designed to solve in the first place. Concerning road networks, the proper accounting of multiple path use to avoid Shannon aggregation error points to an abandonment of Wardrop's equilibrium in favor of Logit choice. This completed shift should favor transit when it is the minority mode.

Key-words: multipath assignment, aggregation of path characteristics, path aggregation, inclusive values, Multinomial Logit, Shannon's measure of information, origins of Random Utility Models (RUM), Probit, Logarithmic Logit, Abraham's Law of traffic assignment, Kirchhoff's distribution, non linearity of Representative Utility Functions (RUF), Box-Cox transformations (BCT), French engineers, Claude Abraham, Stanley Warner, Michel Barbier, Robert Fogel, Daniel McFadden, Abraham-McFadden approach, EOLE, Paris RER E westerly extension, Public Transit (PT) assignment, transit hierarchies, SAMPERS, PRISM, CUBE Voyager, VISUM, NODUS.

Journal of Economic Literature classification: B23, C49, R41.

Table of contents

List of tables

List of figures

1. The necessary aggregation of paths for transport demand model specification

Passenger transport models that explain $D_{od,m}$, the quantity demanded by mode m between any origin o and destination d, require, for all relevant origin-destination (OD) pairs and modes (M), the construction of level-of-service indicators $LOS_{od,m}$ typically comprising at least the modal Cost or Fare $F_{od,m}$ and Travel time $T_{od,m}$ characteristics. Given that other variables ETC are also required to complete specifications, a summary formulation of such demand functions might well be:

$$(0\text{-}A) \qquad D_{od,m} = f_m(LOS_{od,1},...,LOS_{od,M},ETC), \qquad \text{m= 1, ..., M; o, d = 1, ..., Z.}$$

In the particular case of stochastic (random utility) mode choice models commonly treated as demand models, such as the Multinomial Logit (MNL) and Probit, these variables appear in "representative utility functions" (RUF) typically specified only in terms of own-mode LOS characteristics, such as:

$$(0\text{-}B) \qquad V_{od,m} = f_m(LOS_{od,m},ETC), \qquad \text{m= 1, ..., M; o, d = 1, ..., Z.}$$

The construction of such LOS indicators resulting in unique values of modal trip characteristics poses problems of its own because, as a rule, multiple paths of varying "lengths" are in fact used between any given origin and destination: the actual multiplicity of paths has to be both modeled and adequately represented to obtain the unique LOS vector of element values for each OD pair and mode in question. The first task of assignment algorithms therefore consists in espousing traveler choices in networks consisting in sets of techniques[1] or groups of small plants jointly used to simultaneously produce different (trip) outputs. Their second task is to derive the LOS elements from the results of such multiple path assignments where itineraries actually used differ with respect to those elements (time, fare, *etc.*): the question is then how to weigh, or aggregate, them.

In this paper, we assume that the first task has been carried out to perfection and we concentrate on the second to discuss some current weighing practices and their implications for the specification and estimation of the demand functions and their consequent use in project appraisal. The immediate problem giving rise to this concern is for us the extension of a suburban train line westward from central Paris in an urban environment where other competing suburban train lines already exist along with many other transit options (metros, buses, tramways, *etc.*). In the presence of a significant density of transit alternatives, it is tempting for analysts to choose in the menu of transit assignment procedures offered by commercial computer packages the "average LOS" option to generate the transit indicators needed in the demand or mode choice model: we compare such "average" options with the "Logsum" also available from Logit path assignments in some transit assignment programs.

We provide proof of a **P**ath **A**ggregation **TH**eorem (PATH) stating that, in transport networks where unique measures of the utility of multiple paths are required for demand model formulation purposes, constructs based on weighted averages of network path characteristics derived from multipath assignments always underestimate the utility of multiple paths by an amount exactly equal to Shannon's (1948) measure of information if the true path choice model is Multinomial Logit. The aggregation of path characteristics therefore differs from the aggregation of paths in measurable ways, a number of which are also considered in passing.

We study the properties of this measure of aggregation error. First, the issue of proper service aggregation is not limited to public transit projects with closely competing more or less "common" lines in dense urban networks (our particular Paris predicament): it also arises in other modes whenever distinct lines compete within a single mode, and even in freight assignments very briefly alluded to. Notably, the validity of the PATH does not depend on the mathematical form of the path utility functions or on the identification of a common path constant in assignment models.

---

[1] The application of this Hicksian terminology to transport networks was proposed by Åke Andersson at the International Symposium on Travel Supply Models, Montreal, November 17-19, 1977.

2. Does the aggregation of paths differ from that of their characteristics?

2.1 The Logit context and three average constructs of path characteristics

To establish the structural properties of various path weighing methods, we provisionally neglect all observational subscripts — to be reintroduced in the next section where individual observations need to be identified — and reinterpret the remaining running index as applying to paths instead of modes.

If the choice function among path alternatives i ($i = 1, …, M$) is assumed to be MNL, namely:

$$\text{(1-A)} \qquad\qquad p_i = \exp(V_i) / \sum_{i=1}^{M} \exp(V_i) ,$$

$$\text{(1-B)} \qquad\qquad \log p_i = V_i - \ln \sum_{i=1}^{M} \exp(V_i) ,$$

where the *Logsum* or *Inclusive value* term derived by Williams (1977) or McFadden (1978) is easily recognizable. And we wish to consider three ways of aggregating itinerary use by performing a calculation of mean path utility. The first two are readily found in most assignment program menus:

$$\text{(1-C)} \qquad\qquad \bar{V}_p = \sum_i p_i \cdot V_i \qquad\qquad \text{[probabilistic mean]}$$

$$\text{(1-D)} \qquad\qquad \bar{V}_a = \sum_i m_i \cdot V_i \qquad\qquad \text{[arithmetic mean]}$$

where, by convention[2] :

- the p_i denote shares or choice probabilities of the M paths (or itineraries) used ;
- the m_i are all equal to 1/M ;
- the V_i denote, as in (0-B), the representative utility functions (RUF) of the paths.

The third construct, included here for good measure, is inspired by prospect theory which introduces a rupture or "twisting" in the evaluation of choice probabilities, in this case:

$$\text{(1-E)} \qquad \bar{V}_{pp} = \sum_{k=1}^{M} \left[\left(\sum_{i=1}^{k} p_i^* \right)^{\gamma} - \left(\sum_{i=1}^{k-1} p_i^* \right)^{\gamma} \right] \cdot V_k^* \qquad\qquad \text{[prospect power mean]}$$

where :

- starred values of probabilities ($p^*_1, ..., p^*_i, ..., p^*_M$) and of RUF V_k^* signify that the latter are ordered in increasing fashion. And we imagine, in this hypothetical case of a mean construct incorporating an ***attitude towards path utility***, that $\gamma < 1$ stands for instance for the risk of agoraphilia (the fear of little used itineraries), $\gamma > 1$ for the risk of agoraphobia (the fear of heavily used itineraries), and $\gamma = 1$ for a neutral attitude towards itinerary size or inherent attractiveness: in this latter case, (1-E) collapses back to (1-C);

- the simple power transformation p^{γ} is chosen among probability transforming functions that maintain the ***capacity*** of the distribution. Stott (2006, Table 3) lists seven current examples of such functions, obviously excluding from his *ménagerie* the Box & Cox (1964) power transformation applied below to LOS variables but including in it the convoluted inverted *S-shape* animal used by Tversky & Kahneman (1992), $p^{\gamma} / \left(p^{\gamma} + (1-p)^{\gamma} \right)^{1/\gamma}$: all seven do indeed guarantee that transformed probabilities sum to unity.

Our retained *prospect power* formulation of a mean (1-E) is directly borrowed from a model of Air France flight choice between Charles de Gaulle airport and two of the three airports serving London where the author (Lapparent, 2004, 2010) applied the simple power "twist" to Travel time T_k^*, a variable included in the V_i functions of different flights, not to those complete functions themselves. Expressing an attitude towards risk of realization of a RUF instead of a variable, formulation (1-E) is therefore speculative in the sense that it has never been applied as such.

---

[2] The use of these weighted averages appears to be based on intuition or on long-established practice reflected in computer packages: we are unaware of any derivation of (1-A) from utility that would mandate (1-C) or (1-D) as path aggregators.

2.2 Underestimation of the value of paths by constructs averaging their characteristics

We are now in a position to ask how the aggregation of paths differs from that of their characteristics by comparing the *Logsum* to various weighted averages. Generally speaking, consider from (1-B):

$$(2\text{-A}) \qquad V_i = \ln p_i + \ln \left[\sum_i e^{V_i} \right],$$

and, after multiplying all terms by weights w_i normalized to sum to unity, sum over all paths to obtain the weighted mean $\overline{V}_w = \sum_i w_i V_i$, and simply derive:

$$(2\text{-B}) \qquad \overline{V}_w = \sum_i w_i V_i = \sum_i w_i \ln(p_i) + \ln[\sum_j e^{V_j}],$$

from which different special cases will follow, depending on the nature of the weights $\sum_i w_i = 1$. For instance, if the weights are the probabilities themselves as in (1-C), we obtain:

$$(2\text{-C}) \qquad \overline{V}_p - \ln \sum_i \exp(V_i) = \sum_i p_i \cdot \ln(p_i)$$

which can be stated formally as the **p**ath **a**ggregation **th**eorem (PATH):

> *In transport networks where unique measures of the utility of multiple paths are required for demand model formulation purposes, constructs based on probability-weighted averages of network path characteristics derived from multipath assignments always underestimate the utility of multiple paths by an amount equal to Shannon's (1948) measure of information if the true path choice model is Multinomial Logit.*

And we recall for comparison Shannon's own formulation and comment (*op. cit.,* p. 11):

> «*Quantities of the form* $H = -\sum p_i \cdot \ln p_i$ *play a central role in information theory as measures of information, choice, uncertainty. The form of* H *will be recognized as that of entropy in certain formulations of statistical mechanics where p_i is the probability of a system being in cell* i *of its phase space.* H *is then, for example, the* H *in Boltzmann's famous* H *theorem. We shall call* $H = -\sum p_i \cdot \ln p_i$ *the entropy of the set of probabilities* $p_1, \ldots, p_n$.»

Like the entropy measure, the measure $S \equiv \sum_i p_i \cdot \ln(p_i)$ is null if all p_i except one are zero — but this possibility is excluded by the MNL, except if all utilities except one are minus infinity. Otherwise, the properties of the **measure of information S** (or minus-one times the Entropy H) for our path aggregation problem duly match those of the entropy measure, *mutatis mutandis* in view of the sign change. In particular, we note without surprises that *S*:

(i) <u>is negative</u>: the *Logsum* is always larger than (1-C), the *probabilistically weighted average* $\overline{V}_p = \sum_i p_i \cdot V_i$, because probabilities are always positive fractions and all $\ln(p_i) < 0$.

For the same reason, use of the *arithmetic mean* (1-D) yields a difference equal to $\sum_i m_i \cdot \ln(p_i)$, clearly always negative again. In fact, as demonstrated in (2-B), any set of path utility weights normalized to sum to unity will necessarily produce a value inferior to that of the *Logsum*. The difference will be more or less close to *S,* depending on how the weights differ from Shannon's untransformed probabilities: a case in point would be the use of the *prospect power mean* (1-E) leading to a difference equal to $\sum_{k=1}^{M} \left[\left(\sum_{i=1}^{k} p_i^* \right)^{\gamma} - \left(\sum_{i=1}^{k-1} p_i^* \right)^{\gamma} \right] \cdot \ln(p_k^*)$, again negative and $\neq S$ unless $\gamma = 1$;

(ii) <u>is at best constant but generally decreasing with path splitting</u> [S(x,y) ≥ S(x) + S(y)]: the information error concerning a joint path is at least as important as the sum of information errors concerning the paths considered individually.

In the simple case of the splitting of an alternative into two equal options, comparison of S^* [containing the original term $p_c \cdot \ln(p_c)$] to S^{**} [where the original term is replaced by two terms equal to $(p_c/2) \cdot \ln(p_c/2)$], yields $[S^{**} - S^* = -p_c \cdot \ln(2)] < 0$, *ceteris paribus*.

If all options have equal weight, splitting them (or generally increasing their number) will decrease S eventually to 0. Formally, if the measure $S(M)$ pertains to M alternatives of equal probability $1/M$, then $S(M) \rightarrow 0$ as $M \rightarrow \infty$.

(3-A) $$S(M) = M\left[(1/M) \cdot \ln(1/M)\right],$$

(3-B) $$S(M) = \ln(1/M) = -\ln(M) \rightarrow -\infty,$$

(3-C) $$S(M) \rightarrow 0;$$

(iii) <u>has a minimum equal to *ln(M)*</u>, for a constant number of paths M, when all the p_i are equal.

To see this, note that the following minimization problem:

(4-A) $$\underset{p_i}{Min}[\sum_i p_i \ln(p_i)], \text{ s.t. } \sum_i p_i = 1,$$

yields, upon forming the Lagrangian

(4-B) $$L = \sum_i p_i \ln(p_i) + \lambda(\sum_i p_i - 1),$$

the following first order conditions, for all *i*:

(4-C) $$\ln(p_i) + 1 + \lambda = 0$$

implying that all p_i be equal. This means that any change away from equality increases S.

(iv) <u>does not always increase with the variance of probabilities</u>.

Does the fact that changes away from equality of probabilities increase S mean that higher variance in probabilities always increases it as well? To see that it does not, consider firstly particular cases of changes away from equality and secondly the general case:

(a) S increasing with variance: to see that S can increase with the variance of choice[3] probabilities, start with cases of null variance.

Consider first a pair of options of equal importance [$p_1 = (1-p_1) = 0{,}5$]. It is easy to show that we then have $\partial S / \partial p_1 = \ln(p_1) - \ln(1 - p_1) > 0$ *iff* $p_1 > 0{,}5$, namely a measure S that increases with the variance of the choice probabilities: for instance, S increases from -0,6931 to -0,5623 when the probabilities change from (1/2,1/2) to (1/4, 3/4).

Similarly, start with the triplet of options of equal importance (1/3, 1/3, 1/3) and increase the last to $(\varepsilon,\varepsilon,1\text{-}2\varepsilon)$ by drawing equally from the first two. Again, S increases with the variance of probabilities: it equals -1,0986 when all probabilities are equal and -1,0397 if $\varepsilon = 0{,}25$. But this is another case of changes away from equality. What of the more general case?

(b) Sign of changes in S unrelated to that of changes in variance: to see that there exists no general relationship between the direction of changes in the variance of probabilities and that of changes in S, consider more formally the quantities of interest and their modifications following a marginal change in probabilities[4], namely, in succession:

---

[3] The variance of the choice probabilities is not independent from that of the random term associated with each RUF: a large variance of this random term implies that the systematic part of the utility function, to be formulated explicitly below in our discussion of the seminal Abraham (1961) paper, has a relatively smaller role to play as compared to that of the random term. This implies heteroskedasticity of random errors, or some systematic departures from the common homoskedastic variance $\{\pi^2/6\}$. When the ratio between the mean of the systematic part of the utilities and the random term is extremely high, say infinite, any mean will adequately reproduce the utility of the full set of options: the case is analogous to that of the red bus blue bus paradox.

[4] Such that the probabilities sum to one and the sum of their variations equals zero.

(5-A) $$S = \sum_i p_i \cdot \ln(p_i) \qquad and \qquad dS = \sum_i \ln(p_i)\,dp_i \;;$$

(5-B) $$\sigma^2_{p_i} = \sum_i \left[p_i - \frac{1}{M} \right]^2 = \sum_i p_i^2 - \frac{1}{M} \qquad and \qquad d\sigma^2_{p_i} = 2\sum_i p_i\,dp_i\;.$$

To ask whether the sign of dS can be deduced from the sign of $d\sigma^2_{p_i}$, let us interpret these two quantities as scalar products of M-dimensional vectors: the former as the product of vector A with coordinates p_i by the vector C with coordinates dp_i; the latter as the product of vector B with coordinates $ln(p_i)$ by the same vector C with coordinates dp_i.

Consider now a plane, in Figure 1, defined by vectors A and B, as well as Γ, the projection of vector C on that same plane. It is the case that, as a scalar product, dS is of the same sign as that of the *cosine* of the angle formed by A and Γ; for a similar reason, $d\sigma^2_{p_i}$ is of the same sign as that of the *cosine* of the angle formed by B and Γ.

Figure 1. Representation of total differentials of Equations (5-A) and (5-B)
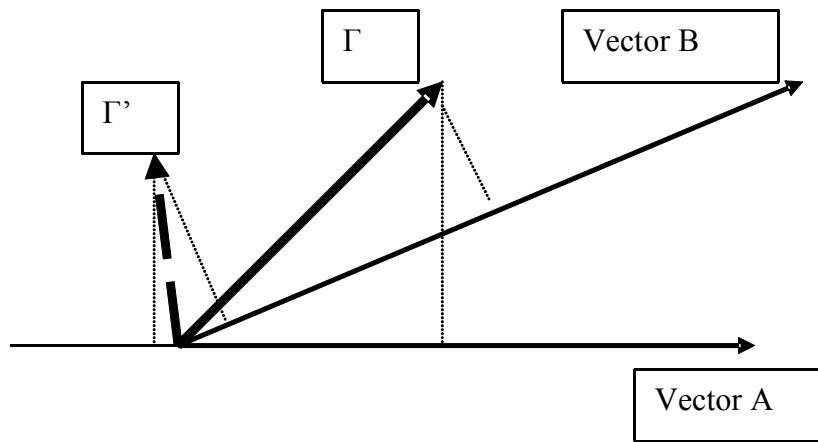
Note that, if A and B are not collinear, one readily finds a vector such as Γ' for which those signs differ and a vector such as Γ for which they are identical: in Figure 1 indeed, the two cosines are of the same sign for Γ but not for Γ'. The signs of the two cosines are identical only if vectors A and B are collinear — the signs of dS and $d\sigma^2_{p_i}$ are then also identical.

Clearly, with elements p_i and $ln(p_i)$, these vectors are obviously not collinear except for very special values of the p_i, namely all equal $1/M$ exemplified in the two particular cases just above in (a). In general, the direction of changes in the variance of probabilities and that of changes in S are unrelated.

(v) is independent from the mathematical form of the utility functions: although obvious from (3-D), this important practical property will shortly be explored at some length in order to search for systematic relationships between the value of S and the mathematical form of the V_i functions.

This matters because Box-Cox transformations introduce asymmetry in the form of the response curves derivable from Box-Cox Logit models. In those models, the marginal utility of the LOS is not assumed to be constant any more, as it is in the popular classical Linear Logit model. The practical question is therefore whether non linearity, which implies the presence of asymmetry of responses to modifications of LOS, affects the aggregation error in analytically predictable ways.

It will turn out that $\Delta S = S_2 - S_1$, the difference between two aggregation errors S_2 and S_1, where indices refer to models varying only in the form of the RUF, requires case by case study: little of an analytic nature can be said about the link between aggregation error and non linearity in the RUF. In fact, rejection by the data of the unreasonable assumption of constant marginal utility does not imply analytical properties of $\Delta S = S_2 - S_1$ that are accessible without use of numerical methods.

2.3 Antecedents for S and consequences of obstinacy in the use of weighted averages

In view of the celebrity of Shannon's information measure, it may be asked whether the notion of path aggregation error *S* is in fact interesting and what benefits could arise by shifting from *Mean* to *Logsum* measures in processing results from path assignment models. Given that people know very well, if only by applying Jensen's inequality, that these measures differ in practice, why not just live with a long-noticed and familiar difference? We argue and show by simulations that recognizing and interpreting this "well known" difference as *S* may prompt a redirection of practice to avoid significant error in the estimation of demand and in the derivation of welfare measures.

2.3.1 Interpretation of a familiar difference: consequences in terms of context

What is in a name? In view of the extreme simplicity of (2-C), is it in fact well known as *S* to all but the present authors? Lest we break down an open door, what are then the published transport demand antecedents of measure *S*? There appears to be two independent streams. In the first, either Shannon's name appears but the formula *S* is used for purposes other than the measurement of aggregation error (2-C) or as the basis for an analogy; similarly, the entropic form -*S* also gives rise to another analogy where the probabilities are replaced by other terms. In the second, *S* is gratuitously computed *ex machina* in the sense of (2-C), but without demonstration and without being named or interpreted after Shannon. At the very least, some benefit may arise from the new interpretation of a "known difference" beyond that of unifying the streams, from which we consider representative cases.

In the first stream, always related to *demand estimation*, one finds for instance Anas (1983) who has shown that the MNL can be derived equivalently by minimization of Shannon's information measure or by maximization of utility. Explicit reference had also been made to Shannon in a Logit context by Theil (1969) who had defined a measure "derived from information theory" for the degree to which a vector P of N expenditure shares p_i, that are each functions of income and of N prices, varies between two periods:

$$(6\text{-}A) \qquad I(p':p) = \sum_i^N p_i' \log p_i' - \log p_i$$

where p_i and p_i' refer to prior and posterior values, and the derived closeness to *S* is obvious[5].

In demand estimation contexts where due reference is made to Shannon's *S*, one also finds Picard (1987) who substitutes flows for probabilities in a minimization of the error between the origin-destination flows D_{ij} to be estimated under various constraints[6] and some *a priori* $\tilde{D}_{ij}$ flows. His objective function, derived from the Kullback-Leibler (1951) distance between distributions[7], is:

$$(6\text{-}B) \qquad MIN \sum_i \sum_j D_{ij} \left[\ln D_{ij} - \ln \tilde{D}_{ij} \right].$$

A closely related construct in terms of quantities of goods is found as part of the following utility function for *n* variants of a differentiated product and for commodity 0, a Hicksian composite good:

$$(6\text{-}C) \qquad U = \sum_i^n a_i X_i - \mu \sum_i^n X_i \ln X_i / N + X_0$$

---

[5] Theil (1965, 1966) had a long-standing interest in applications of information theory to economics.
[6] This thesis is summarized in Picard & Gaudry (1998). The problem was to find, for each of 64 categories of freight transported in Canada, the optimal flows among the 67 principal cities (as origins or destinations), with the flows satisfying row and column total constraints of regional matrices for the cities situated in 8 regions (large provinces or groupings of small provinces). The objective function guarantees that the 64 matrices of estimated flows of dimension 67x67 are "near" exogenously provided *a priori* or observed values.
[7] Hagen-Zanker & Jin (2012) use the same distance measure applied to the normalized flows of trip distribution matrices and duly refer to Shannon.

which is assumed to hold if $\sum_i^n X_i = N$ (*i.e.* if the individual commodities sum to the amount N of the differentiated product) and the a_i and μ (the Logit homoscedasticity term) are nonnegative scalars. But the authors of (6-C) merely refer to its second term as "entropy-type" (Anderson *et al.*, 1986, p. 6; 1988, p. 461) or "of an entropic form" expressing the variety-seeking behavior of the representative consumer (Anderson *et al.*, 1992, p. 79): in contrast with Theil and Picard, they ignore Shannon. In cases (6-B) and (6-C), the analogy[8] arises from the form but the constructs in terms of goods are short of Shannon's information or of Boltzmann's entropy proper which both duly require probabilities[9].

What is in a formula? In the second stream, apparently consisting only in articles or computer programs pertaining to *public transit assignment*, the S formula is again "given" — apparently never derived — but without any reference to Shannon's own S or to the related notion of entropy -S.

For instance one finds, in the appendix of an interesting paper on schedule-based transit assignment (Daly, 1999), the exact expression for S, gratuitously presented as the difference between [$\overline{V}_p = \sum_i p_i \cdot V_i$] and the *Logsum*, with the emphasis put on its negativity, duly imputed to $\ln(p_i) < 0$, but S does not seem to already have a meaning, pedigree, name or existence on its own.

The same restraint prevails in documented commercial computer programs performing transit assignment. The closest term-for-term match to (2-C) is found in the Cube Voyager (2008) manual where the three mathematical expressions belong to a menu of variables calculable from effected multi-path assignments. These output variables are respectively called (*op. cit.*, p. 812) *Average Generalized Cost Skim* [$\overline{V}_p$], *Composite Cost Skim* [$\ln \sum_i \exp(V_i)$] and *Value of Choice* [$\sum_i p_i \cdot \ln(p_i)$].

But no reference to Shannon is found in that manual or in those of other popular programs such as EMME/2 or VISUM which has an option called *Utility* to allow computation of output variable $\sum_i \exp(V_i)$ from realized multipath assignments (PTV AG, p. 464). The forthcoming Emme 3 program (INRO, 2010; Florian & Constantin, 2011), should add to the deterministic[10] optimal strategy assignment implemented in EMME/2 a Logit choice option (Florian & Constantin, 2012).

2.3.2 Consequences in terms of expected benefits

Estimation of demand or mode choice functions with *Logsums* instead of *Means* should improve path choice fit as much as, in the past, similar operations have improved fits in demand models requiring some inclusive value of modal services, traditionally taken to be an average or some "optimal" path.

But that is not all: in the appendix mentioned above, Daly makes an important theoretical point concerning probabilistically weighted averages $\overline{V}_p$ of type (1-C). They can reverse the direction of the effect of path improvements implied by a Logit choice because, as he demonstrates, the effect on $\overline{V}_p$ of improving particular path c must obey the following positivity constraint:

(7-A)
$$\left[\partial \overline{V}_p / \partial V_c = p_c (1 + V_c - \overline{V}_p) \right] > 0$$

---

[8] Another analogy based on a multiplicative form is the use of the expression "Gravity Model" for CES utility functions and, even more frequently, for economic trade and transport models based on Activities and Distance or Cost, rather than on Newton's Masses and Distance.

[9] As recognized by Anderson *et al.* in their footnote on entropy (*op. cit.*, p. 79).

[10] The unrealistic idea of the "optimal strategy" (Spiess & Florian, 1989) is that transit users always walk to the stop or station that generates the lowest generalized path cost for them. Solutions (optimal strategy assignments) are ε-sensitive because, in representative urban areas (*e.g.* Stockholm), or if large zones are used, a traveler departing at a given moment typically has more than a unique path to each destination.

namely any impact of a LOS change on the weighted average $\overline{V}_p$ must be of the same sign as that on path V_c. To understand this requirement, note that, starting with (1-C) and remembering (1-A), we have, if V_c varies by dV_c:

(7-B)
$$d\overline{V}_p = p_c dV_c + \sum_{i \neq c} V_i dp_i \; ;$$

but, as it is the case that for $i \neq c$

(7-C)
$$dp_i = -\frac{e^{V_i} e^{V_c}}{(\sum_i e^{V_i})^2} dV_c = -p_i p_c dV_c$$

and that for c

(7-D)
$$dp_c = [\frac{e^{V_i}}{\sum_i e^{V_i}} - \frac{(e^{V_i})^2}{(\sum_i e^{V_i})^2}] dV_c = [p_c - p_c^2] dV_c \; ;$$

it follows that

(7-E)
$$d\overline{V}_p = (p_c - \sum_i V_i p_i p_c + V_c p_c) dV_c = (p_c + p_c V_c - \overline{V}_p p_c) dV_c \, .$$

Daly's implied condition that $V_c - \overline{V}_p > -1$ for both sides of (7-E) to have the same sign will obviously not hold if V_c is **one or more units smaller** than $\overline{V}_p$. Consequently (*op. cit.,* p. 157):

> "[…] *changes in the average can be used as an approximation to changes in the 'Logsum' value – theoretically the correct value for Logit models – only when the changes relative to the initial probabilities are small*",

lest changes in path *c* produce an incoherent result on the indicator $\overline{V}_p$, *i.e.* violate condition (7-A).

But small changes are precisely not what large transit projects or modal improvements are about. For this reason, not only is $\overline{V}_p$ an inadequate substitute for the *Logsum* in general but it can lead to counterintuitive results in the demand or mode choice model precisely when the change considered is relatively important, relative to existing "reference" options, in the project evaluation context[11].

A case in point is presented in Figure 1, which pertains to Paris, where an extension of RER train line E is envisaged. Over a significant part of its itinerary in the East, this line currently competes with other train lines A and C. But V_E, the current utility of line E in the area affected by the planned westerly extension, is quite low relative to $\overline{V}_{Train}$ because it does not yet exist in the West.

Figure 2. Extension of Regional Express train line E in the densely served centre of Paris

---
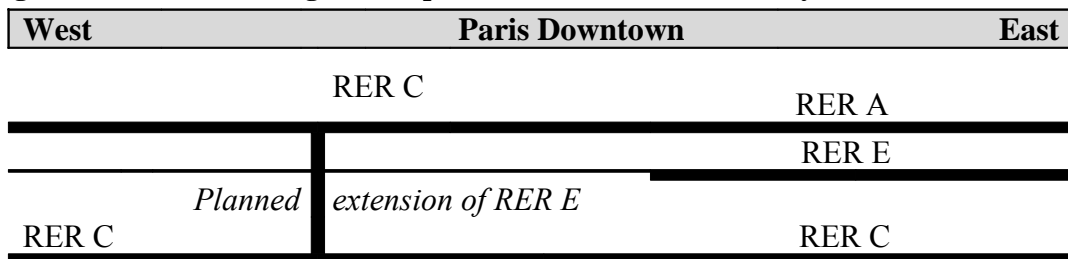
[11] The same problem arises in the hypothetical situation of the addition of a new very bad service *s* with a very low V_s. In this case, the *Logsum* will, by MNL logic, increase despite the fact that $\overline{V}_p$ falls.

The westerly extension, including a new intersection with train line C, would naturally raise V_E considerably, at least relative to the mean. In this case then, use of $\overline{V}_{Train}$ to represent the utility of "trains" in the demand or mode choice functions is likely to violate Daly's condition (7-A).

In fact the benefit is necessarily understated, even if the direction of the effect is not reversed. It is therefore interesting to perform simulations in order to determine, for various representative situations met in urban mode choice models, the impact of using weighted averages rather than *Logsums*.

2.3.3 Consequences in terms of impact on representative minority public transit share

How much does it matter that Shannon aggregation error be avoided in the construction of LOS variables for mode choice models? To get an idea of its importance for Greater Paris, we construct an example from representative utilities assumed known for 3 alternatives: Car with a utility V_1 and Public transit (PT), composed of a first public mode with utility V_{21} and a second one with utility V_{22}.

For two distinct models of Car and PT market shares, index *L* denoting the *Logsum* formulation and index *W* the *Weighted Average* specification of transit utility, we need the four expressions:

$$(8\text{-}A) \quad p_{1L} = \frac{e^{V_1}}{e^{V_1} + e^{\ln\left[e^{V_{21}} + e^{V_{22}}\right]}}, p_{2L} = \frac{e^{\ln\left[e^{V_{21}} + e^{V_{22}}\right]}}{e^{V_1} + e^{\ln\left[e^{V_{21}} + e^{V_{22}}\right]}}; \quad p_{1W} = \frac{e^{V_1}}{e^{V_1} + e^{p_{21}V_{21} + p_{22}V_{22}}}, p_{2W} = \frac{e^{p_{21}V_{21} + p_{22}V_{22}}}{e^{V_1} + e^{p_{21}V_{21} + p_{22}V_{22}}},$$

and we purport to simulate the impact of changes in V_{22} on the difference between PT shares:

$$(8\text{-}B) \quad \Delta p_2 \equiv p_{2L} - p_{2W} = \frac{e^{\ln\left[e^{V_{21}} + e^{V_{22}}\right]}}{e^{V_1} + e^{\ln\left[e^{V_{21}} + e^{V_{22}}\right]}} - \frac{e^{\left(\frac{e^{V_{21}}}{e^{V_{21}} + e^{V_{22}}}\right)V_{21} + \left(\frac{e^{V_{22}}}{e^{V_{21}} + e^{V_{22}}}\right)V_{22}}}{e^{V_1} + e^{\left(\frac{e^{V_{21}}}{e^{V_{21}} + e^{V_{22}}}\right)V_{21} + \left(\frac{e^{V_{22}}}{e^{V_{21}} + e^{V_{22}}}\right)V_{22}}}.$$

In Case A of Table 1, where the actual 2006 Île-de-France[12] daily trip shares of the car (0,64) and transit (0,36) are first reproduced by a *Logsum* term based on a set of hypothesized representative utility values of [2,25; 1,00; 1,00], one finds that substitution of the *Weighted Average* term (smaller by S = -0,6931) underestimates the true transit share by 64%.

Table 1. Impact of Shannon aggregation error on simulated public transit share in Île-de-France

Case A									
				PT service level			Mode shares		Δp_2 (PT share Logsum)-(PT share W.A.) / (PT share W.A.)/100
Alternatives	Vi	Exp(Vi)	pi	Logsum	W. Average		Logsum	W. Average	
Car	2,25	9,48773584	0,64			Car	0,64	0,78	
PT1	1	2,71828183	0,18						
PT2	1	2,71828183	0,18	1,69314718	1,00	PT	0,36	0,22	63,57 % Reference Paris 2006
Total		14,9242995	1,00				1,00	1,00	S = -0,6931
Case B									
				PT service level			Mode shares		
Alternatives	Vi	Exp(Vi)	pi	Logsum	W. Average		Logsum	W. Average	
Car	2,25	9,48773584	0,60			Car	0,60	0,75	
PT1	1	2,71828183	0,17						
PT2	1,25	3,49034296	0,22	1,82593942	1,14	PT	0,40	0,25	59,51 % Increasing PT2 utility by 25%
Total		15,6963606	1,00				1,00	1,00	S = -0,6854
Case C									
				PT service level			Mode shares		
Alternatives	Vi	Exp(Vi)	pi	Logsum	W. Average		Logsum	W. Average	
Car	2,25	9,48773584	0,57			Car	0,57	0,72	
PT1	1	2,71828183	0,16						
PT2	1,5	4,48168907	0,27	1,97407698	1,31	PT	0,43	0,28	53,46 % Increasing PT2 utility by 50%
Total		16,6877067	1,00				1,00	1,00	S = -0,6628

---

[12] The Île-de-France region, comprising the Greater Paris area, had a population of 11,7 million in 2008. The information on shares was graciously provided by *Syndicat des transports d'Île-de-France* (STIF). We neglect here other modes.

Moreover, if the utility of the second public mode is successively increased by 25% in Case B and by 50% in Case C due to assumed service level improvements, PT share forecasts derived from the *Weighted Average* measure of inclusive value continue to underestimate the same shares as forecasted by the *Logsum* measure. But we note that, as *S* increases, the relative superiority of the *Logsum* measure decreases, a property that can be analyzed more formally.

Rewriting (2-C) as $V_L = \bar{V}_p - S$, the ratio of PT shares obtained under each specification in (8-A) is:

(8-C)
$$\frac{p_{2L}}{p_{2W}} = \frac{e^{\bar{V}_p - S}}{e^{V_1} + e^{\bar{V}_p - S}} \mathbf{g} \frac{e^{V_1} + e^{\bar{V}_p}}{e^{\bar{V}_p}} = \frac{e^{V_1 + \bar{V}_p - S} + e^{V_L + \bar{V}_p}}{e^{V_1 + \bar{V}_p} + e^{V_L + \bar{V}_p}}$$

where it is clear that, for a given attractiveness of the other modes represented by V_1, the ratio falls with increases in *S*. It is also clear that, if these other modes have overwhelming market shares, the relative importance of *S* decreases: if the other mode has about two thirds of the market, as in our Paris example, the correction matters.

But, if Shannon path aggregation error can clearly make an important practical difference to demand (and derived welfare) estimation, does the mathematical form of the utility function — important, as we shall recall, since the beginning of RUM path choice modeling in 1961 — mitigate or worsen the error?

3. Shannon's information measure *S* and the functional form of RUF

The mathematical form of random utility functions in transport demand analysis, ignored in the PATH derivation of the *S* measure above, is in fact a fundamental behavioral issue. Linearity is extremely rare in nature and in practice rejected almost every time it is tested in RUF, as a survey of more than 50 passenger and freight mode choice models where both Time and Fare variables were both subjected to Box-Cox transformations, has recently shown for models developed by some 30 researchers belonging to 10 countries (Gaudry, 2011, Tables 7, 8, 9 and 18). As path choice does not differ fundamentally from mode choice, non linearity of Logit path utility functions is also expected and indeed allowed for by many commercial assignment packages[13]: we will now recall how early path and mode choice studies, concerned with form, conscientiously retained non linear formulations.

3.1 The issue of curvature

Most specifications of LOS variables used by Logit practitioners are in fact nested special cases of the Box-Cox transformation (BCT) applicable to any strictly positive variable Var_v:

(8-D)
$$Var_v^{(\lambda)} \equiv \begin{cases} \dfrac{(Var_v)^\lambda - 1}{\lambda} & , \quad \lambda \neq 0, \\ \ln(Var_v) & , \quad \lambda \to 0. \end{cases}$$

and notably to the variables of interest for transport project appraisal, primarily Time (for passengers) and Fare (for freight), present in the RUF (0-B) which can be rewritten explicitly with BCT:

(8-E)
$$V_i = \beta_{i0} + \sum_k \beta_{ik} X_{ik}^{(\lambda_{ik})}$$

As already mentioned above, non linearity, as illustrated in Figure 3 for the binomial case, means that the reaction curve to improvements in variable X_1 associated with alternative 1 will be asymmetric with respect to its inflexion point: it would be symmetric with an inflexion point at $p_1 = 0,50$ only if the data supported in (8-E) the unlikely assumption of constant marginal utility $\lambda_{ik} = 1$, for $\forall i, k$.

Figure 3. Classical Linear-Logit *vs* Standard Box-Cox-Logit Responses
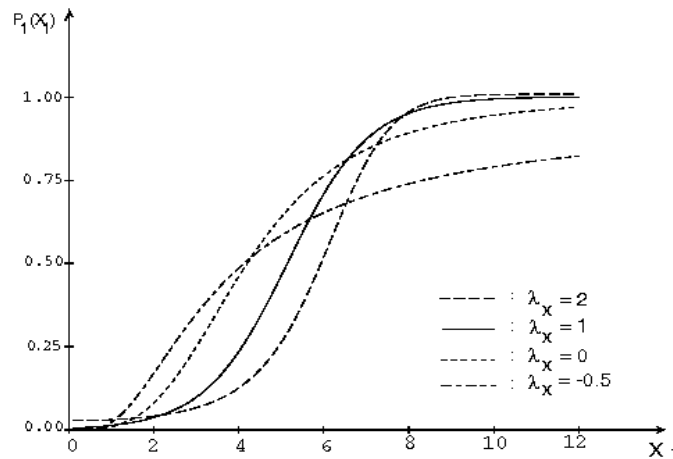
Asymmetry is therefore critically important given that, in forecasts of important changes in LOS, everything is in the curvature because there is no real disagreement on the identity of important variables and because LOS changes considered are far from marginal, consisting for instance in a

---

[13] For instance, VISUM 11.5 (PTV AG, *op. cit.*, p. 507) allows for various options including two where the RUM, consisting solely of a generalized cost, is either transformed logarithmically — in which case Kirchhoff's distribution formula (Fellendorf & Vortisch, 2010, Equation 2.5) is equivalent to Abraham's Law described below —, or subjected to a Box-Cox transformation. If values for the latter are supplied from outside, the invariance problems raised by the absence of intercepts (Schlesselman, 1971) in the path utilities are dodged.

division by two of travel time. In fact, the asymmetric logarithmic response, implying a curve situated above that of the linear response for $[1 < X_1 < 5,5]$ in the case illustrated in Figure 3, prevailed in the careful empiricism of the founders of path and mode choice analysis, as we now recall before formally addressing the issue of the impact of the form of path RUF on the measure S.

3.2 The secret origin of random utility models (RUM) and their functional form

We studied the foundations of random utility modeling, apparently first developed to explain road path choice, and not mode choice, in a close-knit engineering *milieu* of the type described by Ekelund & Hébert (1999) for the "secret origins of microeconomics": we carefully consulted written sources and queried witnesses, notably experienced transport engineers from public and private institutions.

In a nutshell, it seems that formal derivations of formulae based on the Gaussian and Rectangular distributions, published in a scientific journal for roads and airports (Abraham, 1961)[14], served for the *milieu* of French engineers as an explicit justification of prevailing pre-existing road path choice assignment models and of their descendants, all then based on a Logit core.

Notably, a particular variant, sometimes called a "Logarithmic Logit" (*e.g.* Leurent, 1999), was soon designated in official French documents, and remains so to this day, under the name ***Abraham's Law of traffic assignment***, despite the fact that the Logit form itself had not been derived by Claude Abraham who claimed then and now that he is not the author of the so-called "*Loi d'Abraham*" used as an effective approximation of the presumably real McCoys derived and documented at length in his demonstrations. The attribution of this label by the *milieu* is all the less surprising that Abraham had made it crystal clear that it did not really matter[15] what the underlying distribution really is.

For transit assignment problems, the preference went to a linear RUF: Barbier's binomial Probit application to bus and metro, published in 1966, may well have inspired later generalizations by Domencich and McFadden (1975), still all formulated with linear RUF. Now for the "petite histoire".

3.2.1 Logit practice and Abraham's 1961 derivations from Normal & Rectangular distributions

The mood on the banks of the Seine: Logit road choice applications, with non linear utilities, in search of a justification. It is useful to comment on four Parisian steps, distinguished and summarized in Table 2, to document the first RUM:

1) both Setec (1959) and Abraham (1961) provide a panorama of American and French road assignment practices. In the former case of a Channel Tunnel study, Logit forms dominated and were applied with both linear (model M_1) and logarithmic (model M_2) RUF, both of which contained explicit AGC and were considered "approximations" of the Probit, called "*Modèle Normal*";

2-3) all formal derivations from assumed distributions of random terms found in Equation (9-B) below, published or not and based on the Gaussian or on the Rectangular distribution, used linear LOS terms but insisted that they applied as well to logarithms. They were also explicitly derived as "justifications" of then-current Logit practices.

In a footnote (*op. cit.* p. v) of the anonymous Setec derivation, "Mr. Malcor" was credited with the idea of distributed differences in "subjective valuations" but no reference was given. And the formulation is not mentioned in his just-published article on road traffic and operations research (Malcor, 1958).

In his seminal paper, Abraham (1961) considered an individual n choosing between two road paths 1 and 2 with generalized costs (GC) composed of mean linearly weighted Length and Time elements, respectively GC^1 and GC^2. He formulated (*op. cit.,* p. 66)[16] the choice probability, or proportion of users, for the first path as given by:

---

[14] The article, which contains a short presentation by Coquand, Abraham's boss and Director of Roads and Traffic of the French Ministry of transport, is sometimes referenced with this second name, but we follow Frank Haight's usage in his extensive annotated "bibliography in road traffic" (1964) and neglect the author of the administrative *imprimatur*.

[15] To quote him at the end of the section providing the derivation based on the Rectangular distribution: "*Que la distribution réelle des estimations des usagers soit ou non gaussienne, nous n'en savons rien, et cela n'a, au demeurant, pas grande importance*" (*op. cit.*, p. 68).

(9-A) $$p_1 = \text{Prob}\left[U_n^1 - U_n^2 < 0\right],$$

where the utilities of the paths are assumed to be

(9-B) $$U\ _n^1 = GC^1 + \varepsilon_n^1 + \eta_n^1 \text{ and } U\ _n^2 = GC^2 + \varepsilon_n^2 + \eta_n^2,$$

and the two errors of zero mean are respectively associated with Length (Cost) and Time elements of the representative GC term. He then considered how assumptions concerning the distribution of ε and η affected the structure of the path choice model, first deriving a Probit under the assumption of Normal distributions and another model under the alternate assumption of Rectangular distributions. He performed simulations with both structures for 2-path and 3-path cases based on California expressway data and finally addressed related problems, in particular how to handle cases of paths sharing a link and how to calculate revenues from tolls applied only to subsets of links.

Table 2. Four steps in the development of random utility path assignment models around 1961

Given name	Constant in the RUF	Form and variables in the RUF	Source
1. Existing Binomial Logit practice *circa* 1958			**Setec, 1959**
M_1 model	$[\beta_{02} - \beta_{01} = \beta_0]$	$[\beta \cdot Cost_i]$	paragraph 2.220 p. iii
M_2 model		$[\beta \cdot \ln(Cost_i)]$	paragraph 2.221 p. iv
2. Anonymous derivation of the Multinomial Probit, based on the Gaussian distribution, in 1959			**Setec, 1959**
Modèle Normal	$\left[\beta_{0i} = \beta_{0j}, \forall i, j\ paths\right]$	$[\beta \cdot Cost_i]$, or $[\beta \cdot \ln(Cost_i)]$	paragraph 2.60, p. ix
3. Binomial derivations[1] and tests based on Gaussian and Rectangular distributions[2] in 1961			**Abraham, 1961**
Gaussian	$\left[V_{2n} - V_{1n} = \beta_0\right]$	$[\beta_T \cdot Time_i + \beta_C \cdot Cost_i]$	pp. 65-66
Rectangular			pp. 67-68
4. Designation of M_2 Multinomial Logit variant as Abraham's Law after 1961			**Oral tradition & administrative documents**
Loi d'Abraham	$\left[\beta_{0i} = \beta_{0j} = 0, \forall i, j\ paths\right]$	$[\beta \cdot \ln(Generalized\ Cost_i)]$	

[1] The derivation uses a linear utility function but the author considers non linearity much more credible[17].
[2] With standard errors assumed equal across distributions.

Although his formulation (9-A)-(9-B) is identical to that found later in CRA (1972, Ch. 5) or in Domencich & McFadden (1975, Ch. 4, S. 4)[18], he did not consider the Weibull and Cauchy distributions;

4) in particular, the Probit derivation was seen as a justification of Logit practice. We could not successfully date exactly the first use of the expression "Abraham's Law" which combined (1-A) with logarithms of the GC_i to implement non linear path RUF without constants[19]:

(9-C) $$p_i = \exp(V_i) / \sum_{i=1}^{M} \exp(V_i), \text{ with } V_i = \beta \cdot \ln GC_i.$$

In step, French road manuals have long recommended values of β as high as 8 or 10 for intercity road path choice modeled according to "Abraham's Law".

The use of a logarithmic RUF was also voluntary in Warner's (1962) binary urban Logit mode choice model: concerned with goodness-of-fit, he compared various LOS forms[20] and retained the logarithmic one after a very careful analysis of residuals, noticing the inferior fit obtained under a linear form and referring explicitly to traditional log-linear CES production functions.

This is not to say that all path choice models then used in France were straightforward applications of Abraham's Law. But the Probit rapidly became part of the common toolbox: in an analysis of the

---

[16] He claimed in a footnote linked to Equation 9-A that Setec (1959) had first formalized this model in Channel studies. That unpublished consulting report, easily downloadable from the referenced Ministry of Transport site, indeed contains a derivation of a Probit model, with RUF based solely on path costs, under the assumption of "Laplace-Gauss" errors.
[17] As reiterated by Claude Abraham in an email to the authors, dated August 18, 2011: "In the classical [Linear] Logit, there is no difference between a two-minute gain on a 10 minute and on a 60 minute trip, a manifestly absurd hypothesis".
[18] Those authors called their model derived from a Rectangular distribution the "truncated linear probability model".
[19] In practice, this is compensated by correction factors called « bonuses », for instance for highways, *etc*.
[20] Thomas (1967) applies the same careful methodology in a study of car commuter values of time.

profitability of 1961-1962 road works, the Probit curve appears, without further comment, listed among four diversion curve methods (Abraham & Thédié, 1966, p. 145).

Barbier's casual 1963 Probit application: transit choice, with linear utility. This casual state-of-the-art use of the Probit model is also found in transit studies. In 1963, for instance, Michel Barbier, an urban planning engineer[21] working for the Paris Region Planning Institute (IAURP) studied the choice between bus and metro with a sample collected by Setec and SNCF. In his working paper, Barbier (B., 1963b): (i) carried out a full discriminant analysis to find good combinations of costs, frequency, itinerary time and number of transfers which resulted in the same proportion of travelers to a particular destination using the bus or the metro; (ii) formulated an explicit Binomial Probit model of choice between these means of transport based on a difference in generalized cost (a combination of Cost and Time) expressed in time units D, namely[22]:

> *"Noticing that the users' indifference point towards means of transport, such that 50% of users take the bus [rather than the metro], corresponds to a time difference between them of D=5 minutes, we assume that each user takes the bus if he estimates that $\overline{D} \leq 5$. If P(D) denotes the probability of taking the bus, we then have:*

$$(9\text{-}D) \qquad P(D) = \int_{-\infty}^{5} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(\overline{D}-D)^2}{\sigma^2}} \, d\overline{D}$$

> *which, making the change in variables $x = (\overline{D} - D)/\sigma$, can also be written …[…]."*

a formulation, linear in the generalized cost variable, that may have influenced comparable work carried out in Boston in 1970-1971 by Thomas A. Domencich and Daniel McFadden.

Linking the Seine and the Charles rivers. The reason for pointing to this model as a potential source of their inspiration is that Barbier's working paper and its published version are to some extent summarized and referenced[23] in the pair of key documents resulting from these authors' work.

Strangely, however, this connection appears in the literature review (CRA, 1972, p. 3-9; Domencich & McFadden, 1975, p. 25), rather than in the theory development section, and no mention is made of the Probit analysis (9-D). All that is said is:

> *"A French study [Institut d'aménagement et d'urbanisme de la Région Parisienne (1963)] also carried out a full discriminant analysis to find good combinations of times, costs, and number of transfers which resulted in the same proportion of travelers to a particular destination using transit."*

for which the provided reference in both documents is simply:

> Institut d'aménagement et d'urbanisme de la région parisienne (1963). *Étude de divers facteurs influant sur le choix entre autobus et métropolitain par les usagers des lignes S.N.C.F. de banlieue*. Octobre, 1963. Also published in *Transports Urbains* Vol. 1, no. 5.

whereas an informative recognition of Barbier's application would have required something like:

> B., M. (1963b)[24]. *Étude de divers facteurs influant sur le choix entre autobus et métropolitain par les usagers des lignes S.N.C.F. de banlieue*. Direction des études et recherches, Institut d'aménagement et d'urbanisme de la région parisienne (IAURP). 36 p., octobre.

Also published as:

---

[21] Barbier had graduated in 1959 from the *École Nationale des Ponts et Chaussées* (ENPC) and done graduate work in operations research at the Case Institute of Technology in Cleveland. His paper (B., 1963) does not mention Abraham's article.

[22] Our translation is from the French original of the Working paper (B, *op. cit*., p. 16) but the published version (Barbier, 1966) is identical, as may be verified in Appendix 1 below where Figure 4 is captured from it.

[23] The only reference in French among the 126 listed in the CRA version, or in the list augmented to 243 for the book.

[24] Only the author's initials are used on the front page.

Barbier, M. (1966). Choix entre autobus et métropolitain, pp. 27-40, 51-56. Ch. II et Annexes 1-3 in Barbier, M. et Merlin, P., Choix du moyen de transport par les usagers. *Cahiers*[25] *de l'Institut d'aménagement et d'urbanisme de la région parisienne* (*Cahiers de l'IAURP)*[26] Vol. 4-5, 57 pages, avril.

A "Gang of Four". In any case, some 11 years after Abraham's publication, a consulting study by Charles River Associates Incorporated presented in its theoretical section entitled "stochastic specification and estimation techniques" (CRA, 1972, Ch. 5) four derivations from (9-A)-(9-B), effectively adding to previous Gaussian and Rectangular branches new ones based on Weibull and Cauchy error distribution assumptions, yielding the Logit and Arctan form determinations of binomial choice probabilities. This "gang of four", doubling Abraham's twin achievement, even produced *un embarras de richesse*: multinomial derivations were also presented for the Gaussian and Weibull cases[27] and the computational simplicity of Multinomial Logit estimation naturally emphasized.

Randomness in earlier, less general, approaches. Interestingly, the revised version of the consulting report[28], published as a book (Domencich & McFadden, 1975), starts with a laudatory foreword by Richard E. Quandt who had already noted elsewhere (Quandt, *circa* 1974) the extent to which the authors' work constituted a "more general and sophisticated development of" some models by Anthony J. Blackburn and himself. These developments, treating trade-offs between modal cost and time as random variables[29] in models estimated from aggregate data, are very globally referenced in Ch. 4 ("A theory of Population Travel Demand Behavior") of the book as forming one of two earlier approaches yielding functional forms for the choice probabilities, the other being attributed to Thurstone's (1927) "suggested particular case" (*op. cit.*, p. 53). This under-identified nameless particular case[30] turns out to be a Probit application, just like Barbier's model.

RUF forms in Abraham-McFadden (A-M) derivations and in their environment. Published RUM derivations forming "the Abraham-McFadden approach" (Abraham, 1961; McFadden, 1973) all assumed linear forms for the RUF, as did applications of this A-M approach by Barbier (1963b, 1966), CRA (1973) and Domencich & McFadden (1975).

But the environment was changing: if Blackburn (1966) formulates a linear RUF, many, including Warner (1962), practitioners of Abraham's Law, and Quandt (1968), formulate or test and willfully select multiplicative RUF.

---

[25] In the MIT Libraries' catalog listing all volumes from 1 to 41 (except 28), the long form is used. In these *Cahiers* (ISSN 0020-2207), authors' names are found on the back of the cover page of each article or chapter. As Chapter II of Vol. 4-5 pooled Barbier's working paper (1963b) on the choice between bus and metro with other working papers by Merlin on choices between other modal pairs, the reference "Barbier & Merlin (1966)" would also be an adequate reference.

[26] The conventional short form is used on all IAURP *Cahiers* document covers. The long form is «Cahiers de l'Institut d'aménagement et d'urbanisme de la région parisienne». The sub-headings on the front page of Volume 4-5 [*Transports urbains. Les transports urbains et leurs usagers en région de Paris : 2*] refer to *Cahier* theme series and their sub-divisions.

[27] McFadden (1975, Footnote 14) later attributed the Logit derivation from the Weibull distribution solely to McFadden (1973) where empirical estimates are reproduced with permission from CRA (1972).

[28] Without the interesting Appendix A attempt to derive transport demand functions from specific Quadratic, Log-Linear and Stone-Geary utility functions.

[29] Blackburn's Ph.D. thesis, listed among the *Abstracts of theses* for 1966-1967 at M.I.T., and the paper clearly derived from it (Blackburn, 1966), as well as Quandt's article (Quandt, 1968), are contemporaneous with early work on random coefficients in linear regression (Hildreth & Houck, 1968; Swamy, 1970) and might well have their most appropriate place in that literature recently blooming again in Mixed Logit garb in transportation, after a long dormition (*e.g.* Hensher & Johnson, 1979).

[30] This contrasts with the claims of the consulting report where (on page 5-6) Thurstone (1927) had been credited with the first multinomial formulation of Equation (9), also described (on page 4-4) as a model that is "well known (see, for example, Luce, *Individual Choice Behavior*, Wiley, 1959) and need not be elaborated here".

3.2.2 Fogel's unresolved cliometric question of 1964

By contrast with these practitioners of Abraham's Law or with Warner and Quandt, many at the time indeed avoided hard form test work and relied on fast and easy linearity. For instance, Fogel (1964) was interested in defining a measure of changes in generalized transport costs following the introduction of rail in a 19th Century world where water and rail modes competed and maintained positive market shares. Instead of due changes in a *Logsum*, a notion available only later (Small & Rosen, 1981), he used the difference between water and rail costs (extreme cases of 0 and 100% market shares of rail) to build his principal measure, called alpha[31]: $\beta\left(GC_{rail} - GC_{water}\right)$.

As noted in the survey of 50 models referred to above, such *a priori* linear specifications have since been rejected in numerous Logit freight mode choice cases (*op. cit.*, Table 9) where Box-Cox transformations, *e.g.* $\beta\left(GC_{rail}^{(\lambda_C)} - GC_{water}^{(\lambda_C)}\right)$, were tested for rail, road and water, and notably in the very large European canal project *Seine-Nord Europe* linking the Paris area to Belgium (Setec International *et al.*, 2006).

Fogel found that, as an explanatory variable of "social savings", the difference between linear costs (alpha) was not very significant (Davis, 2000). The cliometric question is then whether a more adequate *Logsum* measure, or even appropriately non linear forms of alpha derived with Warner's due care, would have given less disappointing results than those obtained under unlikely linearity.

3.2.3 The impact of RUF form on *S*

Problem formulation. Non linearity of path choice RUF specifications might be established by trial and error, as in applications of Abraham's Law where the coefficient β is conventionally changed manually according to the road type, or by Warner's *ad hoc* method where simple powers of variables are manipulated to improve fit. To reach beyond such special cases, we consider the more general specification of the RUF formally based on determination by the Box-Cox transformation[32] (8-D). The question in practice is then whether there exists a *systematic link* between the value of the BCT on the key LOS variables found in path RUF and the *S* measure.

To answer it, we reintroduce Abraham's observation subscripts, useful here to isolate specific values of LOS variables with particular properties, and compare two choice models differing only in RUF form. By assumption, we have p_{1in} and p_{2in}, the estimated individual shares or choice probabilities for models 1 and 2, formulated with the same number of variables (to simplify notation without loss of generality) but different constraints on the λ BCT parameters applied to the independent variables of each model, namely:

$$(10\text{-A}) \qquad p_{1in} = \frac{\exp V_{1in}}{\sum_m \exp V_{1mn}}, \quad \text{with } V_{1in} = \beta_{1io} + \sum_k \beta_{1ik} X_{1ikn}^{(\lambda_{1ik})},$$

and

$$(10\text{-B}) \qquad p_{2in} = \frac{\exp V_{2in}}{\sum_m \exp V_{2mn}}, \quad \text{with } V_{2in} = \beta_{2io} + \sum_k \beta_{2ik} X_{2ikn}^{(\lambda_{2ik})},$$

where the indices are $i, m = 1, M$ for alternatives, $n = 1, N$ for observations, and where one associates p_{1in}, V_{1in} and V_{1mn} to model 1 and p_{2in}, V_{2in} and V_{2mn} to model 2. In the representative utility functions

---

[31] To estimate alpha, Fogel focused on a sample of 30 of the 825 potential routes between pairs of cities in the West and East of the United States.

[32] There are of course other ways to introduce non linearity of the RUF. For instance, Palma & Picard (1995) successfully use a cubic form on travel time in a Probit model for the Île-de-France region.

V_{1in} and V_{2in} where $k = 1, K$ denotes independent variables and sets $(\beta_{1i0}, \beta_{1ik}, \lambda_{1ik})$ and $(\beta_{2i0}, \beta_{2ik}, \lambda_{2ik})$ summarize the parameters associated to models 1 and 2, respectively.

Note again that use of BCT in path choice models assumes that an alternative-generic constant (AGC) is used on all M paths of each origin-destination pair because invariance of BCT form estimates to changes in the units of measurement of the variables requires the presence of a regression constant (Schlesselman, 1971). In path choice however, none of the M constants can be set at zero for a reference path (as in the case of modes) because paths have no natural labels. But the required AGC constant, which modifies all path shares by adding a common amount to all RUF and consequently has a modeling role of its own, still has to be estimated[33] if one is to stay away from simple power functions which have uncorrectable problems[34].

Path constants are typically ignored in applications of Abraham's (non linear) Law, as they were in (linear) Logit applications to road *tracé* choice (McFadden, 1968, also 1975 & 1976) and to road path choice (Dial, 1971): in the latter case, vehicles are assigned, between an origin and a destination separated by the shortest length L*, to each path of length L "proportionately to $\exp\left[\beta\left(L^* - L\right)\right]$" (*op. cit.,* p. 91).

This said about (10), the difference of interest for our purposes can be defined compactly as follows:

(11)
$$\Delta S = S_2 - S_1 = \underbrace{\{ p_{2i} \ln p_{2i}}_{D_2} - \underbrace{p_{1i} \ln p_{1i} \}}_{D_1} + \sum_{j \neq i} \{ \underbrace{p_{2j} \ln p_{2j}}_{C_2} - \underbrace{p_{1j} \ln p_{1j} \}}_{C_1},$$

where the terms D_2 and D_1 collected in the first parenthesis are direct (own) terms and the others found in the second parenthesis, C_2 and C_1, are cross terms in the sense that we are interested in effects of changes in a particular variable $X_{1iqn} = X_{2iqn} = X_{iqn}$ (such as Time or Fare) that is common to both models but, in (10), appears only in the own RUF of alternative i.

A first question concerning ΔS might be whether there exists a value of $X_{1iqn} = X_{2iqn} = X_{iqn}$ for which errors of aggregation S_2 and S_1 are equal: to find such a ***crossing point***, we could solve $\Delta S_n = 0$ for X_{iqn}^*, but the result would be of limited interest.

Much more interesting for our purposes should be the existence of a maximum or minimum difference between the two measures obtained by first solving $\partial \Delta S_n / \partial X_{iqn} = 0$ for the ***turning point*** X_{iqn}^{**} and then by determining whether it is a maximum or a minimum by considering the second derivative $\partial^2 \Delta S_n / \partial X_{iqn}^2$, evaluating it at the critical point X_{iqn}^{**}, and finding out how it changes signs when X_{iqn} passes through it.

Having found the expression for the second derivative, one could envisage solving $\partial^2 \Delta S_n / \partial X_{iqn}^2 = 0$ to determine the ***inflexion point*** $\tilde{X}_{iqn}$, but this value would also be of very marginal interest for the central issue, that of the existence of a general impact of distinct RUF forms on S, and in particular of the difference between linear symmetric and non linear asymmetric cases.

A focus on the turning point. We therefore focus the analysis on the derivation and properties of the ***turning point*** X_{iqn}^{**}. We present here the short form of the first and second derivatives of ΔS with respect to X_{iqn} and reserve for appendices, using a longer form wherein probabilities are made

---

[33] An approach to the estimation of alternative-generic path constants for all paths is discussed in Gaudry & Tran (2011).
[34] For instance, in contrast to BCT, simple powers do not maintain the order of the data (Johnston, 1984, p. 63).

explicit, X_{iqn} is isolated, and solutions for it can be readily considered — at least qualitatively—, the demonstration that there is no analytical solution for this turning point X_{iqn}^{**}, which must be found and also signed by numerical methods.

Starting with the first derivative of (11) with respect to X_{iqn}, which may be written in short form:

(12)
$$\frac{\partial \Delta S_n}{\partial X_{iqn}} = \left\{ \{1 + \ln p_{2in}\} \cdot \left[\frac{\partial p_{2in}}{\partial X_{iqn}} \right] - \{1 + \ln p_{1in}\} \cdot \left[\frac{\partial p_{1in}}{\partial X_{iqn}} \right] \right\}$$
$$+ \sum_{j \neq i} \left\{ \{1 + \ln p_{2jn}\} \cdot \left[\frac{\partial p_{2jn}}{\partial X_{iqn}} \right] - \{1 + \ln p_{1jn}\} \cdot \left[\frac{\partial p_{1jn}}{\partial X_{iqn}} \right] \right\},$$

it is made clear in Appendix B, using the long form of (12), that finding the critical value $X_{iqn} = X_{iqn}^{**}$ that equalizes it to zero is not analytically feasible and requires numerical methods.

But of course this value $X_{iqn} = X_{iqn}^{**}$ will be a maximum or a minimum depending on whether the second derivative evaluated at that point is negative or positive. For this sign determination, the second derivative of (11) with respect to X_{iqn} may in turn be written in short form from (12) as:

(13)
$$\frac{\partial^2 \Delta S_n}{\partial X_{iqn}^2} = \left\{ \left[\{1 + \ln p_{2in}\} \left[\frac{\partial^2 p_{2in}}{\partial^2 X_{iqn}} \right] + \frac{1}{p_{2in}} \left[\frac{\partial p_{2in}}{\partial X_{iqn}} \right]^2 \right] - \left[\{1 + \ln p_{1in}\} \left[\frac{\partial^2 p_{1in}}{\partial^2 X_{iqn}} \right] + \frac{1}{p_{1in}} \left[\frac{\partial p_{1in}}{\partial X_{iqn}} \right]^2 \right] \right\}$$
$$+ \sum_{j \neq i} \left\{ \left[\{1 + \ln p_{2jn}\} \left[\frac{\partial^2 p_{2jn}}{\partial^2 X_{iqn}} \right] + \frac{1}{p_{2jn}} \left[\frac{\partial p_{2jn}}{\partial X_{iqn}} \right]^2 \right] - \left[\{1 + \ln p_{1jn}\} \left[\frac{\partial^2 p_{1jn}}{\partial^2 X_{iqn}} \right] + \frac{1}{p_{1jn}} \left[\frac{\partial p_{1jn}}{\partial X_{iqn}} \right]^2 \right] \right\}$$

and it is shown in Appendix C, using the long form of (13) with terms arranged to again put the X_{iqn} in evidence, that such sign determination also requires a numerical exercise.

The maximum or minimum difference between error measures S_2 and S_1 therefore bears a systematic link with the functional form of the RUF but that linkage can only be evaluated by numerical methods, on a case by case basis: it always depends on many variables and parameters in non linear ways unfortunately not amenable to analytic treatment.

This conclusion remains if the comparison between the two models distinguished solely by functional form of the RUF is effected between any sophisticated nonlinear (asymmetric response) Box-Cox Logit and a pedestrian Linear (symmetric response) Logit.

4. How many *Logsums* for dense transit networks?

The clear superiority of *Logsum* over weighted average LOS measures implies that the choice among paths within transit and car networks should be addressed by Logit models. In many actual modal networks, this could mean accounting for subtle differences among transit modes or road types by dummy variables associated to the relevant links of the transit or road networks and concentrating the effort on the explanation of path shares between origin-destination points. This includes a role for path constants, notably when the long-demonstrated non linearity in LOS is handled by Box-Cox transformations rather than by simple power functions of generalized costs[35] or of individual LOS service variables. But modes have specificities.

A baker's combinatorial dozen? In public transit networks, should a compromise layer of branches be considered between the traditional modes (car, transit, on foot, *etc*.) and the transit path access means, especially when the number of transit modes is plethoric? Our Paris predicament is that there are at least 4 different types of buses[36] (Ordinary, Bus Rapid Transit (BRT), T-Zen[37], Local minibuses), 2 kinds of tramways (large ones on rails, with high windows; smaller ones on tires) and of metros (ordinary and automatic) and regional trains of quite different characteristics, "feel" and comfort. If a hierarchy is considered, which of these 10+ means are the high modes and which the low modes merely serving as access to the higher modes and requiring a path access model of their own?

And should the transit hierarchy depend on the direction of a return trip? Conceivable hierarchies[38] are many but are not nested in a statistical sense. It might well be easier to obtain specific cross-effects (including complementarity) by including in the RUF of a MNL the characteristics of some other modes (with due constraints on LOS forms), as in classical microeconomic demand systems, than to decide on the most credible hierarchy of higher and lower (access) modes among 10 transit modes.

Is Wardrop equilibrium in palliative care? Path costs are always generalized costs. So, if equilibrium methods are used to model road path choice, two acute problems arise. First, even in the simplistic case where time and cost intervene linearly, user equilibrium is unique only if users have a single value of time or if cost and time change with flow on each link in identical manner (Dafermos 1983). Moreover, as in Wardrop's equilibrium link flows are unique but the number of itineraries used is unknown and their identity is not analytically derivable from the optimal solution[39], their due aggregation by *Logsums* is problematic. The necessity of identifying all itineraries effectively used, in conformity with the path aggregation theorem (PATH), should prompt a movement of analysts and commercial programs away from equilibrium assignment and towards the use of Logit based methods.

This hold also for extensions of Wardrop equilibrium to public transit passenger flows, as in Pirandello (Piron & Delons, 2007) where volume-delay curves are replaced by volume-discomfort curves ─ with comfort defined by the number of users per square meter of vehicle space. The same objection applies to replacement of comfort by other notions, such as wait-time at stops.

Are freight assignment models exceptions? It might be thought that freight, where choice of mode and path is often combined within an extended mode-and-path abstract choice formulation, would avoid the multipath assignment and aggregation problems discussed above for passengers and still

---

[35] As noted above, the Kirchhoff formula (PTV AG, *op. cit*., p. 511) is indistinguishable from Abraham's Law.

[36] Among the 1 433 bus lines covering 24 660 km of routes, many are complementary with the rail system but many are in competition with it. The metro part of the rail system currently has 300 stations on 215 km of lines and the Grand Paris Automated Metro would add 175 km of lines and about 60 stations.

[37] T-Zen buses in service since 2011 in the Paris area benefit from dedicated Bus Rapid Transit (BRT) lanes but have tramway-type doors and windows. Fish or fowl?

[38] A Bus-Train layer was developed in the first version of the SAMPERS regional model for Sweden (Transek, 1999) and abandoned when the model was revised and updated in 2003 (Transek, 2004a, 2004b). Other models, such as PRISM, developed for Birmingham (Rand Europe, 2004), had many fewer transit modes than Paris.

[39] Sometimes authors use very astute patches (*e.g.* Bar-Gera, 2006; Bar-Gera *et al.*, 2012) to compensate for this lack.

obtain relatively good fits by simple shortest path methods. But that would underestimate the refinement of freight assignment procedures. For instance NODUS, originally conceived in this way (Jourquin, 1995; Jourquin & Beuthe, 2006) has recently added a procedure which retains the least cost mode-path itinerary and assigns the origin-destination flow to competing itineraries in proportion to costs (Jourquin & Limbourg, 2007): but proportions bring us back to Abraham's Law...

It would seem that, between the opposite excesses of non nested hierarchies of dubious meaning and simple-minded shortest path and equilibrium assignments, multiple path realism combined to the *Logsum* can go a long way, notably in plethoric networks, towards making sense of observed multiple path use and reflect the value added by enriching already dense modal networks, urban or not.

5. Conclusion

We have interpreted the often mentioned difference between *Logsum* and average utility in terms of Shannon's (1948) information measure *S*, leading to a **P**ath **A**ggregation **TH**eorem (PATH). It states that, in transport networks where unique measures of the utility of multiple paths are required for demand model formulation purposes and the true path choice model is Multinomial Logit (MNL), constructs based on weighted averages of path characteristics derived from multipath assignments always underestimate the utility of multiple paths, a deficit exactly equal to *S* (corresponding to minus-one times entropy) if the weights are the path choice probabilities.

We have studied the properties of this *S* measure of aggregation error, along with those arising from other types of averages of path characteristics, outlining some implications for demand or mode choice model formulation and project appraisal. Notably, the validity of the PATH does not depend on the specific contents of the representative utility functions (RUF) associated to paths, such as their mathematical form or their eventual inclusion of alternative-generic constants (AGC). We have showed by simulation that averaging modes or sub-modes — a frequent feature of traffic modeling studies — can lead to important errors in terms of level of traffic and of welfare measurement.

Concerning the mathematical form of the RUF, we have recalled that, after the publication of Abraham's 1961 random utility model (RUM) of road path choice deriving the Probit specification based on the Gaussian error distribution (and another specification based on the Rectangular error distribution), French engineers used this seminal approach as justification of road path choice formulae then in current use and assigned the name "Abraham's Law" to a particular standard one, effectively a "Logarithmic Logit" close to the logarithmic RUF carefully specified for Logit mode choice by Warner in 1962. For transit assignment problems, the preference went to a linear RUF: Barbier's casual binomial Probit application to bus and metro, published in 1966, may have inspired the later generalizations by Domencich and McFadden.

In view of many founders' conscientiously crafted nonlinear Logit formulations, and more generally of the repeatedly demonstrated presence of nonlinearity in RUF path and mode specifications since their careful work 50 years ago, we have analyzed the impact of such nonlinearity on *S*. This impact is tractable through a comparison of measures S_2 and S_1 associated with two path choice models differing only in RUF form, as determined by Box-Cox transformations applied to their level-of-service (LOS) variables. We have showed that, although the difference between measures S_2 and S_1 may reach a minimum or a maximum with changes in LOS, the solution for such a turning point cannot be established analytically but requires the use of numerical methods. The demonstrable impact on *S* of nonlinearity, or asymmetry of Logit curve response, is consequently tractable only at non trivial computational cost.

We have pointed out that the path aggregation issue, whereby aggregation of paths by *Logsums* differs from aggregation of their characteristics by averages, is not limited to public transit (PT) projects with more or less "common" lines competing in dense urban transit networks (our particular Paris predicament) but also arises in other modes whenever distinct itineraries or lines compete within a single mode. Concerning dense urban PT networks, we have hypothesized that *Logsums* based on multiple path assignments treating all transit means (about 10 in our problem) as one modal network should, using Ockham's razor, be simpler than the insertion of a layer of choice hierarchies among such urban means based on non nested specifications embodying assumptions on the identity of "higher" and "lower" means, the latter reasserting the multiple path access problems the hierarchies were designed to solve in the first place. Concerning road networks, the proper accounting of multiple path use to avoid Shannon aggregation error points to an abandonment of Wardrop's equilibrium in favor of Logit choice, as it does for any application of Wardrop to transit path choice.

This shift should favor transit when it is the minority mode.

6. Acknowledgements

7. References

Abraham, C. (1961). La répartition du trafic entre itinéraires concurrents: réflexions sur le comportement des usagers, application au calcul des péages. *Revue générale des routes et aérodromes* 357, pages 57-60, 65-72, 75-76.

Abraham, C., Thédié, C. (1966). *Rentabilité des travaux routiers. Cycles d'études 1961-1962. Rapport définitif.* Direction des Routes et de la Circulation Routière, Ministère de l'Équipement, Paris, 167 pages.

Anas, A. (1983). Discrete choice theory, information theory and the multinomial logit and gravity models. *Transportation Research B* 17, 1, 13-23.

Anderson, S., Palma, A. de, Thisse, J.-F. (1986). *A Representative Consumer Theory of the Logit Model.* Discussion Paper 1986002, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE), 13 p., November.

Anderson, S., Palma, A. de, Thisse, J.-F. (1988). A Representative Consumer Theory of the Logit Model, *International Economic Review*, 29, 461-466.

Anderson, S., Palma, A. de, Thisse, J.-F. (1992). *Discrete Choice Theory of Product Differentiation*, MIT Press, 423 p., January.

Blackburn, A. J. (1966). A nonlinear model of passenger demand. In *Studies in Travel Demand*, Vol. II, pp. 47-89, MATHEMATICA, Princeton, N.J. Also in Quandt, R. E. (1970). *The Demand for Travel: Theory and Measurement*, pp. 163-199, Heath Lexington.

B., M. (1963a). *Modèles de trafic. Analyse bibliographique.* Institut d'aménagement et d'urbanisme de la région parisienne (IAURP), 70 p., août.

B., M. (1963b). *Étude de divers facteurs influant sur le choix entre autobus et métropolitain par les usagers des lignes S.N.C.F. de banlieue.* Direction des études et recherches, Institut d'aménagement et d'urbanisme de la région parisienne (IAURP). 36 p., octobre.

Barbier, M. (1966b). Choix entre autobus et métropolitain, pp. 27-40, 51-56. Ch. II et Annexes 1-3 in Barbier, M. et Merlin, P., Choix du moyen de transport par les usagers. *Cahiers de l'IAURP* No 4-5, 57 pages, avril.

Bar-Gera, H. (2006). Primal Method for Determining the Most Likely Route Flows in Large Networks. *Transportation Science* 40, 3, 269-286.

Bar-Gera, H., Boyce, D., Nie, Y. M. (2012). User-equilibrium route flows and the condition of proportionality. *Transportation Research B* 46, 3, 440-462, March.

Box, G. E. P., Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society B*, 26, 211-243.

CRA (1972). *A disaggregated behavioural model of urban travel demand*. Report No. CRA-156-2 prepared under contract No. FH-11-7566 "Development of a behavioural modal split model" for the Federal Highway Administration. Charles River Associates Incorporated, Cambridge, Massachusetts, March.

Cube Voyager (2008). *Reference Guide Version 5.0*. Document revision 50-006-1, 1000 pages, Citilabs, 12 December.

Dafermos, S. (1983). A Multicriteria Route-Mode Choice Traffic Equilibrium Model. *Bulletin of the Greek Mathematical Society* 24, 13-32.

Daly, A. (1999). The use of schedule-based assignments in public transport modeling. *Proceedings of 27th European Transportation Forum*, Seminar F, Cambridge, England, 149-157.

Davis, L. (2000). *Review Essay*. Economic History Association. EH.net, June 30.

Domencich, T. A., McFadden, D. (1975). *Urban Travel Demand: A Behavioral Analysis*, North Holland, Amsterdam, 215 p.

Dial, R. B. (1971). A Probabilistic Multipath Traffic Assignment Model which Obviates Path Enumeration. *Transportation Research* 5, 83-111.

Ekelund, R. B., Hébert, R. F. (1999). *Secret Origins of Modern Microeconomics: Dupuit and the Engineers*. 468 pages, The University of Chicago Press.

Fellendorf, M., Vortisch, P. (2010). Microscopic Traffic Flow Simulator VISSIM. Ch. 2 in Barceló, J. (ed.), *Fundamentals of Traffic Simulation*. International Series in Operations Research & Management Science 145, pages 63-93. Springer Verlag.

Florian, M., Constantin, I. (2011). *Emme strategy transit assignments with variants.* INRO, Montreal, 87 pages, 5th May. www.inro.ca/en/index.php.

Florian, M., Constantin, I. (2012). A note on logit choices in strategy transit assignment. *European Journal of Transportation and Logistics* 1,1, 18 pages, forthcoming.

Fogel, R. W. (1964). *Railroads and American Economic Growth: Essays in Econometric History*, Baltimore: Johns Hopkins Press, 296 pages.

Gaudry, M., Jara-Diaz, S. R., Ortuzar, J. de D. (1989). Value of Time Sensitivity to Model Specification. *Transportation Research B* 23, 2, 151-158.

Gaudry, M. (2011). *Quebec-Windsor Corridor High Speed Rail Market Forecast Profiles in Context: Level-of-Service Response Curvature Sensitivity and Attitude to Risk or to Distance in Forty Logit Core Model Applications of the Law of Demand*. 100 pages, February. www.e-ajd.net.

Gaudry, M., Quinet, E. (2011). *Agréger des caractéristiques ou des chemins? Annexe 1 de la Note sur la problématique de l'évaluation de l'utilité économique d'Éole: dix problèmes et quelques propositions de remèdes.* Communication présentée à la séance plénière du Comité scientifique d'évaluation économique du projet de prolongement du RER E à l'ouest (EOLE), Réseau Ferré de France, Paris, 6 juillet.

Gaudry, M., Tran, C.-L. (2011). *Identifying all alternative-specific constants in Multinomial Logit models by Inverse Power Transformation Capture*. Publication AJD-141, Agora Jules Dupuit, Université de Montréal, 23 pages, juillet. www.e-ajd.net.

Gaudry, M., Briand, A, Paulmyer, I., Tran, C.-L. (2008). Choix modal transpyrénéen ferroviaire, intermodal et routier: un modèle Logit Universel de forme Box-Cox généralisée. In Lebacque, P., Boillot, F. and M. Aron (eds), *Modélisation du trafic – Actes du groupe de travail 2004-2005*, Actes INRETS N^{0}110, p.11-119, décembre. www.e-ajd.net.

Hagen-Zanker, A., Jin, Y. (2012). A New Method of Adaptive Zoning for Spatial Interaction Models. Department of Architecture, University of Cambridge, *Geographical Analysis*, forthcoming.

Haight, F. A. (1964). An annotated bibliography of scientific research in road traffic and safety. *Operations Research* 12, 6, 976-1039, December.

Hensher, D. A. and L. W. Johnson (1979). A Random Coefficient Model of the Determinants of Frequency of Shopping Trips. *Australian Economic Papers* 18, 33, pages 322-336, December.

Hildreth, C., Houck. J. P. (1968). Some estimators for a linear model with random coefficients. *Journal of the American Statistical Association* 63, 584-595.

INRO (2010). Emme 3 at www.inrosoftware.com .

Johnston, J. (1984). *Econometric Methods*, Third Edition, McGraw-Hill Book Company.

Jourquin, B. (1995). *Un outil d'analyse économique des transports de marchandises sur des réseaux multi-modaux et multi-produit. Le réseau virtuel, concept, méthodes et applications.* Thèse de Ph.D., Facultés Universitaires Catholiques de Mons.

Jourquin, B., Beuthe, M. (2006). A decade of freight transport modelling with virtual networks: Acquired experiences and new challenges. GTM-paper, April 2004, in A. Reggiani and P. Nijkamp (eds.), *Spatial Dynamics, Networks and Modelling*, Edward Elgar, 2006.

Jourquin B., Limbourg, S. (2007). Une procédure d'affectation multi-flux et multimodale appliquée aux réseaux transeuropéens de transport de fret. *Les cahiers scientifiques du transport*, 9-26.

Kullback, S., Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics* 22, 1, 79-86.

Lapparent, M. de (2004). *La valeur du temps dans les déplacements réguliers professionnels*. Thèse d'économie, Université Paris I–Panthéon-Sorbonne. Publication AJD-81, Agora Jules Dupuit, Université de Montréal, 180 p., juin. www.e-ajd.net.

Lapparent, M. de (2010). Attitude towards Risk of Time Loss in Travel Activity and Air Route Choices. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, forthcoming. Revised version of: Optimism, Pessimism toward Risk of Time Losses in Business Air Travel Demand, and Insurance Premium. Publication AJD-77, Agora Jules Dupuit, Université de Montréal, 29 p., (www.e-ajd.net).

Leurent, F. (1999). *Calcul de deux modèles d'affectation*. Rapport de convention SETRA-INRETS, Référence INRETS: 18-98 SETRA, 41 p., février.

Luce, D. C. (1959, 2005). *Individual Choice Behavior: A Theoretical Analysis*, John Wiley & Sons, Inc., New York, 1959. Also *Individual Choice Behavior: A Theoretical Analysis,* 153 pages, Dover Publications, Mineola, New York, 2005.

Malcor, R. (1958). Théorie de la circulation routière et recherche opérationnelle. *Travaux* 280, p. 137-142, février.

McFadden, D. (1968). *The Revealed Preferences of a Government Bureaucracy*. Technical Report No 17, Institute of International Studies, University of California at Berkeley, 52 p.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behaviour. Ch. 4 in Zaremka, P. (ed.). *Frontiers in Econometrics*, pp. 105-142, Academic Press, New York.

McFadden, D. (1975). The Revealed Preferences of a Government Bureaucracy: Theory. *Bell Journal of Economics* 6, 401-416.

McFadden, D. (1976). The Revealed Preferences of a Government Bureaucracy: Empirical Evidence, *Bell Journal of Economics* 7, 55-72.

McFadden, D. (1978). Modelling the Choice of Residential Location. In Karlvist, A., Lundqvist, A., Snickars, F., Weibull, J. (eds), *Spatial Interaction Theory and Planning Models*. North-Holland, 75-96.

Palma, A. de., Picard, N. (1995). Route choice decisions under uncertainty. *Transportation Research A* 39, 295-324.

Piron, V., Delons, J. (2007). Pirandello, un nouvel outil d'aide à la décision. *Transports* 449, p. 5-15, mai-juin.

PTV AG. *PTV Vision: VISUM 11.5 – Bases théoriques*, 812 pages.

Quandt. R. E. (1968). Estimation of Modal Splits. *Transportation Research* 2, 41-50. Also in Quandt, R. E. (1970). *The Demand for Travel: Theory and Measurement*, pp. 147-162, Heath Lexington.

Quandt, R. E. (c. 1974). *Basic Model Types in Estimating Transport Demand*. Mimeo, 11 p., Economics Department, Dickinson Hall, Princeton University, Princeton, N.J.

Rand Europe (2004). *PRISM West Midlands: Tour based mode destination modelling*, 196 pages, 6th July. http://www.prism-wm.com/.

Schlesselman, J. (1971). Power families: a note on the Box and Cox transformation. *Journal of the Royal Statistical Society* Series B 33, 307-31.

Setec (1959). *Tunnel sous la Manche – Étude de Trafic et de Rentabilité: Partie B – Annexe*. The Economist Intelligence Unit Ltd (Londres), De Leuw, Cather and Cᵒ (Chicago) et Société d'Études Techniques et Économiques (Paris), pages iii-ix (pages 196-202, Volume 3), novembre. http://temis.documentation.developpement-durable.gouv.fr/document.xsp?id=Temis-0005494&qid=sdx_q0&n=1&q=.

Setec International, Stratec, Eurotrans (2006). *Évaluation socio-économique du projet SNE*. Canal Seine-Nord Europe: Études d'avant-projet. Voies Navigables de France, Béthune, Document STP 1700, 35 pages, 27 janvier.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal* 27, pp. 379-423 & 623-656, July & October.

Small, K. A., Rosen, H. S. (1981). Applied Welfare Economics with Discrete Choice Models. *Econometrica* 49, 3, January.

Spiess, H., Florian, M. (1989) Optimal strategies: a new assignment model for transit networks. *Transportation Research B* 23, 2, 83–102.

Stott, H. P. (2006). Cumulative prospect theory's functional menagerie. *Journal of Risk Uncertainty* 32, 101–130.

Swamy, P. A. V. B. (1970). Efficient Inference in a Random Coefficient Regression Model. *Econometrica* 38, 311-323.

Theil, H. (1965). The information approach to demand analysis. *Econometrica* 33, 67-87, January.

Theil, H. (1966). *Economics and Information Theory*. Vol. 7 in the series "Studies in Mathematical and Managerial Economics", North Holland.

Theil, H. (1969). A Multinomial Extension of the Linear Logit Model. *International Economic Review* 10, 3, 251-259.

Thomas, T. C. (1967). *The value of time for passenger cars: an experimental study of commuters' values*. Report PB-175-731, Stanford Research Institute, May.

Thurstone, L. (1927). A law of comparative judgement. *Psychological Review* 34, 273-286.

Transek (1999). *The SAMPERS System: Overview*. 25 pages, Transek AB, Solna.

Transek (2004a). *Utveckling av Sampers 2.1. Del 1 Estimeringeninkl tilläggstester*. Report 2004:1 171 pages, Transek AB, Solna, January.

Transek (2004b). *The SAMPERS System 2.1: Overview*. 29 pages, Transek AB, Solna.

Tversky, A., Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5, 297–323.

Warner, S. L. (1962). *Stochastic Choice of Mode in Urban Travel: A Study in Binary Choice*. Northwestern University Press, 90 pages.

Williams, H. C. W. L. (1977). On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit. *Environment and Planning* 9a, 3, 285-344.

8. Appendix A. Barbier's Probit transit mode choice model

Figure 4 presents a pasted extract from Barbier (1966, Annexe 3, p. 55), repeating the title of the appendix in which it is found.

Figure 4. *Formulation théorique des résultats observés pour le choix entre autobus et métropolitain*

L'observation des résultats obtenus conduit à émettre l'hypothèse que l'estimation par les usagers de la différence de temps D' — 4 répond à une loi « normale » de moyenne D' — 4 et dont l'écart quadratique moyen est de la forme :

$$\sigma = a + b \left| D' - 4 \right|$$

En remarquant que le point d'indifférence des usagers envers le moyen de transport, 50 % des usagers prenant l'autobus, correspond à D' = 5, nous supposerons que chaque usager choisit l'autobus s'il estime :

$$\overline{D'} \leqslant 5$$

En désignant par P (D') la probabilité pour qu'un usager prenne l'autobus, on a donc :

$$P(D') = \int_{-\infty}^{5} \frac{1}{\sigma \sqrt{2\pi}} \, e^{-\frac{1}{2} \frac{(\overline{D'} - D')^2}{\sigma^2}} \, d\overline{D'}$$

En faisant le changement de variable : $x = \dfrac{\overline{D'} - D'}{\sigma}$

$$P(D') = \int_{-\infty}^{\frac{5 - D'}{\sigma}} \frac{1}{\sqrt{2\pi}} \, e^{-\frac{1}{2} x^2} = \Pi \left(\frac{5 - D'}{\sigma} \right)$$

D'après la valeur observée de P (D'), on peut donc calculer des valeurs empiriques σ (D') et chercher à estimer les coefficients a et b par la méthode des moindres carrés.

On trouve ainsi :

$$\sigma = 3{,}93 + 0{,}52 \left| D' - 4 \right|$$

On a alors calculé la courbe théorique :

$$P(D') = \Pi \left[\frac{5 - D'}{3{,}93 + 0{,}52 \, (D' - 4)} \right]$$

9. Appendix B. The long form of the first derivative

We wish to document more explicitly, but still partially[40] for brevity, the steps taken to isolate X_{iqn} in (B-2), the long form of (12), the first derivative of ΔS with respect to X_{iqn}.

Making probabilities in (12) explicit and first isolating X_{iqn} only in all partial derivatives, one obtains:

(B-1)
$$\frac{\partial \Delta S_n}{\partial X_{iqn}} = \left\{ 1 + V_{2in} - \ln \sum_m \exp V_{2mn} \right\} \odot \left[\frac{\beta_{2iq} X_{iqn}^{\lambda_{2iq}-1}}{\frac{\sum_{m \neq i} \exp V_{2mn}}{\exp V_{2in}} + \frac{\exp V_{2in}}{\sum_{m \neq i} \exp V_{2mn}} + 2} \right]$$

$$- \left\{ 1 + V_{1in} - \ln \sum_m \exp V_{1mn} \right\} \odot \left[\frac{\beta_{1iq} X_{iqn}^{\lambda_{1iq}-1}}{\frac{\sum_{m \neq i} \exp V_{1mn}}{\exp V_{1in}} + \frac{\exp V_{1in}}{\sum_{m \neq i} \exp V_{1mn}} + 2} \right]$$

$$+ \sum_{j \neq i} \left\{ \begin{array}{l} -\left\{ 1 + V_{2jn} - \ln \sum_m \exp V_{2mn} \right\} \odot \left[\frac{\exp V_{2jn}}{\left(1 + \sum_{m \neq i} \exp V_{2mn} / \exp V_{2in} \right)^2} \beta_{2iq} X_{iqn}^{\lambda_{2iq}-1} \right] \\ + \left\{ 1 + V_{1jn} - \ln \sum_m \exp V_{1mn} \right\} \odot \left[\frac{\exp V_{1jn}}{\left(1 + \sum_{m \neq i} \exp V_{1mn} / \exp V_{1in} \right)^2} \beta_{1iq} X_{iqn}^{\lambda_{1iq}-1} \right] \end{array} \right\}$$

However, as $\exp V_{1in} = \exp V_{1in}^* \exp(\beta_{1iq} X_{iqn}^{\lambda_{1iq}} / \lambda_{1iq})$ and $\exp V_{2in} = \exp V_{2in}^* \exp(\beta_{2iq} X_{iqn}^{\lambda_{2iq}} / \lambda_{2iq})$, where $V_{1in}^* = \beta_{1i0} + \sum_{k \neq q} \beta_{1ik} X_{1ikn}^{(\lambda_{1ik})} - (\beta_{1iq} / \lambda_{1iq})$ and $V_{2in}^* = \beta_{2i0} + \sum_{k \neq q} \beta_{2ik} X_{2ikn}^{(\lambda_{2ik})} - (\beta_{2iq} / \lambda_{2iq})$, equation (B-1) can be rewritten in a form where X_{iqn} no more appears in the V_{1mn} or V_{2mn} of the direct terms, in the V_{1jn} or V_{2jn} of the cross terms, or in the V_{1in}^* and V_{2in}^* belonging to both.

Concerning these terms, note in passing that, if $\lambda_{1iq} = \lambda_{2iq} = 0$, the $-(\beta_{1iq} / \lambda_{1iq})$ and $-(\beta_{2iq} / \lambda_{2iq})$ terms, now absent from V_{1in}^* and V_{2in}^*, are reassigned inside $\exp(\beta_{1iq} X_{iqn}^{\lambda_{1iq}} / \lambda_{1iq})$ and $\exp(\beta_{2iq} X_{iqn}^{\lambda_{2iq}} / \lambda_{2iq})$ to yield $\exp(\beta_{1iq} \log X_{iqn})$ and $\exp(\beta_{2iq} \log X_{iqn})$.

Now, assuming that $\lambda_{1iq}, \lambda_{2iq} \neq 0$, we obtain the desired long form that is clearly without analytical solution for the desired turning point $X_{iqn} = X_{iqn}^{**}$:

---

[40] The detailed steps taken from (12) to (B-2) for all first derivatives are described in full in Gaudry *et al.* (2008).

$$\frac{\partial \Delta S_n}{\partial X_{iqn}} = \left\{ 1 + \left[V_{2in}^* + (\beta_{2iq} X_{iqn}^{\lambda_{2iq}} / \lambda_{2iq}) \right] - \ln \left[\exp V_{2in}^* \exp(\beta_{2iq} X_{iqn}^{\lambda_{2iq}} / \lambda_{2iq}) + \sum_{m \neq i} \exp V_{2mn} \right] \right\} \subseteq$$

$$\left[\frac{\sum_{m \neq i} \exp V_{2mn}}{\exp V_{2in}^* \exp(\beta_{2iq} X_{iqn}^{\lambda_{2iq}} / \lambda_{2iq})} + \frac{\exp V_{2in}^* \exp(\beta_{2iq} X_{iqn}^{\lambda_{2iq}} / \lambda_{2iq})}{\sum_{m \neq i} \exp V_{2mn}} + 2 \right]^{-1} \beta_{2iq} X_{iqn}^{\lambda_{2iq}-1}$$

$$- \left\{ 1 + \left[V_{1in}^* + (\beta_{1iq} X_{iqn}^{\lambda_{1iq}} / \lambda_{1iq}) \right] - \ln \left[\exp V_{1in}^* \exp(\beta_{1iq} X_{iqn}^{\lambda_{1iq}} / \lambda_{1iq}) + \sum_{m \neq i} \exp V_{1mn} \right] \right\} \mathbf{g}$$

$$\left[\frac{\sum_{m \neq i} \exp V_{1mn}}{\exp V_{1in}^* \exp(\beta_{1iq} X_{iqn}^{\lambda_{1iq}} / \lambda_{1iq})} + \frac{\exp V_{1in}^* \exp(\beta_{1iq} X_{iqn}^{\lambda_{1iq}} / \lambda_{1iq})}{\sum_{m \neq i} \exp V_{1mn}} + 2 \right]^{-1} \beta_{1iq} X_{iqn}^{\lambda_{1iq}-1}$$

(B-2)
$$+ \sum_{j \neq i} \left\{ \begin{array}{l} - \left\{ 1 + \left[V_{2jn}^* \right] - \ln \left[\exp V_{2in}^* \exp(\beta_{2iq} X_{iqn}^{\lambda_{2iq}} / \lambda_{2iq}) + \sum_{m \neq i} \exp V_{2mn} \right] \right\} \mathbf{g} \\ \left[\dfrac{\exp V_{2jn}}{\left[1 + \exp\left(-\beta_{2iq} X_{iqn}^{\lambda_{2iq}} / \lambda_{2iq}\right) \sum_{m \neq i} \exp\left(V_{2mn} - V_{2in}^*\right) \right]^2} \right] \beta_{2iq} X_{iqn}^{\lambda_{2iq}-1} \\ + \left\{ 1 + \left[V_{1jn}^* \right] - \ln \left[\exp V_{1in}^* \exp(\beta_{1iq} X_{iqn}^{\lambda_{1iq}} / \lambda_{1iq}) + \sum_{m \neq i} \exp V_{1mn} \right] \right\} \mathbf{g} \\ \left[\dfrac{\exp V_{1jn}}{\left[1 + \exp\left(-\beta_{1iq} X_{iqn}^{\lambda_{1iq}} / \lambda_{1iq}\right) \sum_{m \neq i} \exp\left(V_{1mn} - V_{1in}^*\right) \right]^2} \right] \beta_{1iq} X_{iqn}^{\lambda_{1iq}-1} \end{array} \right\},$$

.

where

$\lambda_{1iq}, \lambda_{2iq} \neq 0$, $V_{2in}^* = \beta_{2i0} + \sum_{k \neq q} \beta_{2ik} X_{2ikn}^{(\lambda_{2ik})} - (\beta_{2iq} / \lambda_{2iq})$ and $V_{1in}^* = \beta_{1i0} + \sum_{k \neq q} \beta_{1ik} X_{1ikn}^{(\lambda_{1ik})} - (\beta_{1iq} / \lambda_{1iq})$.

10. Appendix C. The long form of the second derivative

We wish to document more explicitly, but still partially[41] for brevity, the intermediate steps taken to isolate X_{iqn} in (C-6), the long form of the second derivatives of ΔS with respect to X_{iqn}.

For this, we focus on the first and third components of (13) corresponding to direct term D_{2n} and cross term C_{2n} in (11). The second component, D_{1n}, may be obtained from the first by replacing p_{2in} with p_{1in} and the fourth component, C_{2n}, from the third by replacing p_{2jn} with p_{1jn}.

In component D_{2n}, the first and second derivatives of p_{2in} with respect to X_{iqn} are, in turn:

$$
\begin{aligned}
\frac{\partial p_{2in}}{\partial X_{iqn}} &= \frac{\partial}{\partial X_{iqn}}\left(\frac{\exp V_{2in}}{\exp V_{2in}+\sum_{m\neq i}\exp V_{2mn}}\right)\\
&= p_{2in}\left(\frac{1}{\exp V_{2in}}\frac{\partial \exp V_{2in}}{\partial X_{iqn}}-\frac{1}{\exp V_{2in}+\sum_{m\neq i}\exp V_{2mn}}\frac{\partial \exp V_{2in}}{\partial X_{iqn}}\right)\\
\text{(C-1)}\qquad &= p_{2in}\left(\frac{1}{\exp V_{2in}}-\frac{1}{\exp V_{2in}+\sum_{m\neq i}\exp V_{2mn}}\right)\exp V_{2in}\frac{\partial V_{2in}}{\partial X_{iqn}}\\
&= p_{2in}\left(1-\frac{\exp V_{2in}}{\exp V_{2in}+\sum_{m\neq i}\exp V_{2mn}}\right)\beta_{2iq}X_{iqn}^{\lambda_{2iq}-1}\\
&= p_{2in}\left(1-p_{2in}\right)\beta_{2iq}X_{iqn}^{\lambda_{2iq}-1}
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial^2 p_{2in}}{\partial X_{iqn}^2} &= \frac{\partial}{\partial X_{iqn}}\left(\frac{\partial p_{2in}}{\partial X_{iqn}}\right)=\frac{\partial}{\partial X_{iqn}}\left[p_{2in}\left(1-p_{2in}\right)\beta_{2iq}X_{iqn}^{\lambda_{2iq}-1}\right]\\
&= \frac{\partial p_{2in}}{\partial X_{iqn}}\left[\left(1-p_{2in}\right)\beta_{2iq}X_{iqn}^{\lambda_{2iq}-1}\right]+p_{2in}\frac{\partial}{\partial X_{iqn}}\left[\left(1-p_{2in}\right)\beta_{2iq}X_{iqn}^{\lambda_{2iq}-1}\right]\\
\text{(C-2)}\qquad &= \frac{\partial p_{2in}}{\partial X_{iqn}}\left[\left(1-p_{2in}\right)\beta_{2iq}X_{iqn}^{\lambda_{2iq}-1}\right]+p_{2in}\left[\frac{\partial\left(1-p_{2in}\right)}{\partial X_{iqn}}\beta_{2iq}X_{iqn}^{\lambda_{2iq}-1}+\left(1-p_{2in}\right)\frac{\partial\left(\beta_{2iq}X_{iqn}^{\lambda_{2iq}-1}\right)}{\partial X_{iqn}}\right]\\
&= \frac{\partial p_{2in}}{\partial X_{iqn}}\left[\left(1-p_{2in}\right)\beta_{2iq}X_{iqn}^{\lambda_{2iq}-1}\right]+p_{2in}\left[-\frac{\partial p_{2in}}{\partial X_{iqn}}\beta_{2iq}X_{iqn}^{\lambda_{2iq}-1}+\left(1-p_{2in}\right)\beta_{2iq}\left(\lambda_{2iq}-1\right)X_{iqn}^{\lambda_{2iq}-2}\right],
\end{aligned}
$$

where replacing $\partial p_{2in}/\partial X_{iqn}$ with its value given in (C-1) allows for the following formulation:

$$
\begin{aligned}
\text{(C-3)}\qquad &= p_{2in}\left(1-p_{2in}\right)^2\beta_{2iq}^2 X_{iqn}^{2(\lambda_{2iq}-1)}+p_{2in}\left[-p_{2in}\left(1-p_{2in}\right)\beta_{2iq}^2 X_{iqn}^{2(\lambda_{2iq}-1)}+\left(1-p_{2in}\right)\beta_{2iq}\left(\lambda_{2iq}-1\right)X_{iqn}^{\lambda_{2iq}-2}\right]\\
&= p_{2in}\left(1-p_{2in}\right)^2\beta_{2iq}^2 X_{iqn}^{2(\lambda_{2iq}-1)}-p_{2in}^2\left(1-p_{2in}\right)\beta_{2iq}^2 X_{iqn}^{2(\lambda_{2iq}-1)}+p_{2in}\left(1-p_{2in}\right)\beta_{2iq}\left(\lambda_{2iq}-1\right)X_{iqn}^{\lambda_{2iq}-2}\\
&= p_{2in}\left(1-p_{2in}\right)\beta_{2iq}X_{iqn}^{\lambda_{2iq}-1}\left[\left(1-p_{2in}\right)\beta_{2iq}X_{iqn}^{\lambda_{2iq}-1}-p_{2in}\beta_{2iq}X_{iqn}^{\lambda_{2iq}-1}+\left(\lambda_{2iq}-1\right)X_{iqn}^{-1}\right]\\
&= p_{2in}\left(1-p_{2in}\right)\beta_{2iq}X_{iqn}^{\lambda_{2iq}-1}\left[\left(1-2p_{2in}\right)\beta_{2iq}X_{iqn}^{\lambda_{2iq}-1}+\left(\lambda_{2iq}-1\right)X_{iqn}^{-1}\right],
\end{aligned}
$$

whereby replacing $\partial p_{2in}/\partial X_{iqn}$ and $\partial^2 p_{2in}/\partial X_{iqn}^2$ by these values now yields for component D_{2n}

---

[41] Further documentation of the detailed steps of these manipulations from (13) to (6-C) is available from the authors.

$$= (1 + \ln p_{2in}) p_{2in} (1 - p_{2in}) \beta_{2iq} X_{iqn}^{\lambda_{2iq}-1} \left[(1 - 2p_{2in}) \beta_{2iq} X_{iqn}^{\lambda_{2iq}-1} + (\lambda_{2iq} - 1) X_{iqn}^{-1} \right] + \frac{1}{p_{2in}} p_{2in}^2 (1 - p_{2in})^2 \beta_{2iq}^2 X_{iqn}^{2(\lambda_{2iq}-1)}$$

$$= (1 + \ln p_{2in}) p_{2in} (1 - p_{2in}) \beta_{2iq} X_{iqn}^{\lambda_{2iq}-1} \left[\left(1 - 2p_{2in} + \frac{1 - p_{2in}}{1 + \ln p_{2in}} \right) \beta_{2iq} X_{iqn}^{\lambda_{2iq}-1} + (\lambda_{2iq} - 1) X_{iqn}^{-1} \right]$$

$$= p_{2in} (1 - p_{2in}) \beta_{2iq} X_{iqn}^{\lambda_{2iq}-1} \left\{ \left[(1 - 2p_{2in})(1 + \ln p_{2in}) + 1 - p_{2in} \right] \beta_{2iq} X_{iqn}^{\lambda_{2iq}-1} + (1 + \ln p_{2in})(\lambda_{2iq} - 1) X_{iqn}^{-1} \right\}$$

$$= p_{2in} (1 - p_{2in}) \beta_{2iq}^2 X_{iqn}^{2(\lambda_{2iq}-1)} \left\{ \left[(1 - 2p_{2in})(1 + \ln p_{2in}) + 1 - p_{2in} \right] + (1 + \ln p_{2in}) \frac{\lambda_{2iq} - 1}{\beta_{2iq} X_{iqn}^{\lambda_{2iq}}} \right\}$$

$$= p_{2in} (1 - p_{2in}) \beta_{2iq}^2 X_{iqn}^{2(\lambda_{2iq}-1)} \left\{ 1 - p_{2in} + (1 + \ln p_{2in}) \left[1 - 2p_{2in} + \frac{\lambda_{2iq} - 1}{\beta_{2iq} X_{iqn}^{\lambda_{2iq}}} \right] \right\}.$$

In component C_{2n}, by contrast, since only V_{2in}, which includes X_{iqn}, appears in $\sum_{m \neq j} \exp V_{2mn}$, the first and second derivatives of p_{2jn} with respect to X_{iqn} are, in turn:

(C-4)
$$\frac{\partial p_{2jn}}{\partial X_{iqn}} = \frac{\partial}{\partial X_{iqn}} \left(\frac{\exp V_{2jn}}{\exp V_{2jn} + \sum_{m \neq j} \exp V_{2mn}} \right) = \frac{\partial}{\partial X_{iqn}} \left(\frac{1}{1 + \sum_{m \neq j} \exp(V_{2mn} - V_{2jn})} \right), (j \neq i)$$

$$= -\frac{1}{\left[1 + \sum_{m \neq j} \exp(V_{2mn} - V_{2jn}) \right]^2} \frac{\partial \left[1 + \sum_{m \neq j} \exp(V_{2mn} - V_{2jn}) \right]}{\partial X_{iqn}}$$

$$= -\frac{1}{\left[1 + \sum_{m \neq j} \exp(V_{2mn} - V_{2jn}) \right]^2} \frac{\partial \left(\sum_{m \neq j} \exp V_{2mn} \right)}{\partial X_{iqn}}$$

$$= -\frac{1}{\left[1 + \sum_{m \neq j} \exp(V_{2mn} - V_{2jn}) \right]^2} (\exp V_{2in}) \beta_{2iq} X_{iqn}^{\lambda_{2iq}-1}$$

$$= -p_{2jn}^2 (\exp V_{2in}) \beta_{2iq} X_{iqn}^{\lambda_{2iq}-1}$$

and

(C-5)
$$\frac{\partial^2 p_{2jn}}{\partial X_{iqn}^2} = \frac{\partial}{\partial X_{iqn}} \left(\frac{\partial p_{2jn}}{\partial X_{iqn}} \right) = \frac{\partial \left[-p_{2jn}^2 (\exp V_{2in}) \beta_{2iq} X_{iqn}^{\lambda_{2iq}-1} \right]}{\partial X_{iqn}} \quad (j \neq i)$$

$$= -\frac{\partial p_{2jn}^2}{\partial X_{iqn}} (\exp V_{2in}) \beta_{2iq} X_{iqn}^{\lambda_{2iq}-1} - p_{2jn}^2 \frac{\partial \left[(\exp V_{2in}) \beta_{2iq} X_{iqn}^{\lambda_{2iq}-1} \right]}{\partial X_{iqn}}$$

$$= -2 p_{2jn} \frac{\partial p_{2jn}}{\partial X_{iqn}} (\exp V_{2in}) \beta_{2iq} X_{iqn}^{\lambda_{2iq}-1} - p_{2jn}^2 \left[\frac{\partial \exp V_{2in}}{\partial X_{iqn}} \beta_{2iq} X_{iqn}^{\lambda_{2iq}-1} + \exp V_{2in} \frac{\partial \left(\beta_{2iq} X_{iqn}^{\lambda_{2iq}-1} \right)}{\partial X_{iqn}} \right]$$

$$= -2 p_{2jn} \left[-p_{2jn}^2 (\exp V_{2in}) \beta_{2iq} X_{iqn}^{\lambda_{2iq}-1} \right] (\exp V_{2in}) \beta_{2iq} X_{iqn}^{\lambda_{2iq}-1}$$

$$\qquad - p_{2jn}^2 \left[(\exp V_{2in}) \beta_{2iq}^2 X_{iqn}^{2(\lambda_{2iq}-1)} + (\exp V_{2in}) \beta_{2iq} (\lambda_{2iq} - 1) X_{iqn}^{\lambda_{2iq}-2} \right]$$

$$= 2 p_{2jn}^3 (\exp 2V_{2in}) \beta_{2iq}^2 X_{iqn}^{2(\lambda_{2iq}-1)} - p_{2jn}^2 (\exp V_{2in}) \beta_{2iq}^2 X_{iqn}^{2(\lambda_{2iq}-1)} \left(1 + \frac{\lambda_{2iq} - 1}{\beta_{2iq} X_{iqn}^{\lambda_{2iq}}} \right)$$

$$= p_{2jn}^2 (\exp 2V_{2in}) \beta_{2iq}^2 X_{iqn}^{2(\lambda_{2iq}-1)} \left[2 p_{2jn} - \left(1 + \frac{\lambda_{2iq} - 1}{\beta_{2iq} X_{iqn}^{\lambda_{2iq}}} \right) \right]$$

whereby replacing $\partial p_{2jn}/\partial X_{iqn}$ and $\partial^2 p_{2jn}/\partial X_{iqn}^2$ by these values now yields for component C_{2n}

$$= (1+\ln p_{2jn})\left\{p_{2jn}^2(\exp 2V_{2in})\beta_{2iq}^2 X_{iqn}^{2(\lambda_{2iq}-1)}\left[2p_{2jn}-\left(1+\frac{\lambda_{2iq}-1}{\beta_{2iq}X_{iqn}^{\lambda_{2iq}}}\right)\right]\right\}+\frac{1}{p_{2jn}}\left[-p_{2jn}^2(\exp V_{2in})\beta_{2iq}X_{iqn}^{\lambda_{2iq}-1}\right]$$

$$= p_{2jn}^2(\exp 2V_{2in})\beta_{2iq}^2 X_{iqn}^{2(\lambda_{2iq}-1)}\left\{(1+\ln p_{2jn})\left[2p_{2jn}-\left(1+\frac{\lambda_{2iq}-1}{\beta_{2iq}X_{iqn}^{\lambda_{2iq}}}\right)\right]-\frac{p_{2jn}}{\beta_{2iq}X_{iqn}^{\lambda_{2iq}-1}}\right\}.$$

And the long form may now be written with terms arranged to again put the X_{iqn} in evidence:

(C-6)
$$\frac{\partial^2 \Delta S_n}{\partial X_{iqn}^2} = p_{2in}(1-p_{2in})\beta_{2iq}^2 X_{iqn}^{2(\lambda_{2iq}-1)}\left\{1-p_{2in}+(1+\ln p_{2in})\left[1-2p_{2in}+\frac{\lambda_{2iq}-1}{\beta_{2iq}X_{iqn}^{\lambda_{2iq}-2}}\right]\right\}$$

$$-p_{1in}(1-p_{1in})\beta_{1iq}^2 X_{iqn}^{2(\lambda_{1iq}-1)}\left\{1-p_{1in}+(1+\ln p_{1in})\left[1-2p_{1in}+\frac{\lambda_{1iq}-1}{\beta_{1iq}X_{iqn}^{\lambda_{1iq}-2}}\right]\right\}$$

$$+\sum_{j\neq i} p_{2jn}^2\beta_{2iq}^2 X_{iqn}^{2(\lambda_{2iq}-1)}(\exp 2V_{2in})\left\{(1+\ln p_{2jn})\left[2p_{2jn}-\left(1+\frac{\lambda_{2iq}-1}{\beta_{2iq}X_{iqn}^{\lambda_{2iq}}}\right)\right]-\frac{p_{2jn}}{\beta_{2iq}X_{iqn}^{\lambda_{2iq}-1}}\right\}$$

$$-\sum_{j\neq i} p_{1jn}^2\beta_{1iq}^2 X_{iqn}^{2(\lambda_{1iq}-1)}(\exp 2V_{1in})\left\{(1+\ln p_{1jn})\left[2p_{1jn}-\left(1+\frac{\lambda_{1iq}-1}{\beta_{1iq}X_{iqn}^{\lambda_{1iq}}}\right)\right]-\frac{p_{1jn}}{\beta_{1iq}X_{iqn}^{\lambda_{1iq}-1}}\right\}$$

where the four probabilities and their respective RUF are, successively for own terms:

$$p_{2in} = 1\big/\left[1+\sum_{m\neq i}\exp(V_{2mn}-V_{2in})\right]; \qquad p_{1in} = 1\big/\left[1+\sum_{m\neq i}\exp(V_{1mn}-V_{1in})\right],$$

$$V_{2mn} = \beta_{2m0}+\sum_k \beta_{2mk}X_{2mkn}^{(\lambda_{2mk})},\ (m\neq i); \qquad V_{1mn} = \beta_{1m0}+\sum_k \beta_{1mk}X_{1mkn}^{(\lambda_{1mk})},\ (m\neq i),$$

$$V_{2in} = \beta_{2i0}+\beta_{2iq}X_{iqn}^{(\lambda_{2iq})}+\sum_{k\neq q}\beta_{2ik}X_{2ikn}^{(\lambda_{2ik})}; \qquad V_{1in} = \beta_{1i0}+\beta_{1iq}X_{iqn}^{(\lambda_{1iq})}+\sum_{k\neq q}\beta_{1ik}X_{1ikn}^{(\lambda_{1ik})},$$

and for cross terms:

$$p_{2jn} = 1\big/\left[1+\sum_{m\neq j}\exp(V_{2mn}-V_{2jn})\right],\ (j\neq i); \qquad p_{1jn} = 1\big/\left[1+\sum_{m\neq j}\exp(V_{1mn}-V_{1jn})\right],\ (j\neq i),$$

$$V_{2mn} = \begin{cases}\beta_{2m0}+\sum_k \beta_{2mk}X_{2mkn}^{(\lambda_{2mk})},\ (m\neq i)\\ \beta_{2i0}+\beta_{2iq}X_{iqn}^{(\lambda_{2iq})}+\sum_{k\neq q}\beta_{2ik}X_{2ikn}^{(\lambda_{2ik})},\ (m=i)\end{cases}; \ V_{1mn} = \begin{cases}\beta_{1m0}+\sum_k \beta_{1mk}X_{1mkn}^{(\lambda_{1mk})},\ (m\neq i)\\ \beta_{1i0}+\beta_{1iq}X_{iqn}^{(\lambda_{1iq})}+\sum_{k\neq q}\beta_{1ik}X_{1ikn}^{(\lambda_{1ik})},\ (m=i)\end{cases},$$

$$V_{2jn} = \beta_{2j0}+\sum_k \beta_{2jk}X_{2jkn}^{(\lambda_{2jk})},\ (j\neq i); \qquad V_{1jn} = \beta_{1j0}+\sum_k \beta_{1jk}X_{1jkn}^{(\lambda_{1jk})},\ (j\neq i).$$

an expression that clearly cannot be signed analytically for given turning point values $X_{iqn} = X_{iqn}^{**}$.