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

by

Marc Gaudry*

	1	Directeur, Agora Jules Dupuit (AJD), www.e-ajd.net
		Professeur honoraire, Département de sciences économiques
		Université de Montréal
		Montréal, Canada
	2	Chercheur associé, www.inrets.fr
		Institut National de Recherche sur les Transports et leur Sécurité (INRETS)
		Université Paris-Est
		Noisy-le-Grand, France

Université de Montréal
Agora Jules Dupuit, Publication AJD-127

Université Paris-Est
Institut National de Recherche sur les Transports et leur Sécurité
Département économie et sociologie des transports, Working Paper DEST N° 12

Version 1 on July 23, 2008; Version 2 of April 30, 2014

0.1. Abstract, key words and JEL classification

Setting a context for a discussion of rail demand function curvatures. As envisaged High Speed Rail (HSR) projects typically anticipate a reduction by half or more of existing rail Travel Time, ridership and revenue forecasts depend critically on the curvature of the rail demand curve defined with respect to both Fare and Time levels of service (LOS), in addition to Frequency. We discuss these curvature issues and derived forecasts within the Quasi-Direct Format (QDF) architecture in use since the US Northeast Corridor Transportation Project of the 1960's: a product of Total Market size and Mode split component models.

The Logit component in a QDF structure admitting power transformations of variables. We study curvatures with Box-Cox transformations (BCT) applied to the variables of both components but our discussion focuses on the Logit piece used in representative QDF structures because modal diversion effects of new HSR services tend to dominate the slower long-term induction effects of increased market size, so that net discounted project results are driven by Linear Logit specifications prevailing to this day, in transport as in other applications.

The establishment of BCT non linearity and its use in forecasts. Our analysis starts with a reference summary of results from three Canadian mode choice models formulated in 1976-1978 and in 1992-1994 for the purposes of forecasting the effects of major infrastructure changes, respectively new airports in Southern Ontario and faster rail in the Quebec-Windsor Corridor. We emphasize how HSR revenue maximization of rail Fares and Speeds obtained under hypothesized Linear forms then yielded higher revenues than under data determined optimal Box-Cox forms even if their market shares were higher on relatively long distances. The rest of the paper interprets or positions these results among others and further documents Corridor model forecasts.

The *prima facie* meaning of gross BCT profiles in 40 models. Concerning interpretation, we point out that, although basic consumer demand theory does not constrain admissible values of BCT in Total demand and Logit Mode choice components of Modal demand models, actual estimates are in fact generally compatible with "Cost damping" claims: (a) Time and Cost sensitivities (expressed as first partial derivatives of the demand function) typically fall with Distance in passenger and freight markets, except in urban passenger markets where Time sensitivity almost always increases; (ii) relative sensitivities (the Value of Time) always increase with Distance, irrespective of whether slopes fall at a decreasing (damped) or at an increasing (amplified) rate, in the 40-some surveyed models built by some 30 researchers for 10 countries. Such empirical regularities establish a fundamental difference between intercity and urban markets based on the second derivatives of demand functions.

Gross and net power values. We further suggest that such real gross BCT power value sensitivity profiles estimated without taking the attitude to risk into account could in fact reflect two effects that, as shown in recent seminal work on Rank Dependent Utility (RDU), can be identified by products of power functions of Fare or Time, to wit a simple power to determine *the attitude to outcome risk (probability)* and a BCT power to determine the *attitude to outcome proper*. We also reinterpret recent models making successful use of interactions between a Distance variable raised to a simple power and a LOS variable (especially Travel Time) sometimes raised to a BCT power as identifying, also by a product of functions, an "attitude to Distance" that allows for an explicit breakdown of gross attitudinal parameters between *attitude to distance* and *attitude to outcome* elements.

Robustness of results from estimated curvatures. We imply overall that, beyond mere fit and other demonstrated benefits, untested linear forms of Standard Logit utility function variables are theoretically unexpected as representations of price-time utility maps, statistically unsustainable in many samples ranges where *prima facie* gross cost or time damping or amplification prevail in absolute and relative senses, practically biased as conditional bargaining games to play with data, and often demonstrably unsound or misleading in the production of HSR passenger and revenue forecasts, and no doubt elsewhere as well.

Key words: *Logit, Box-Cox transformation, Box-Tukey transformation, High Speed Rail, Canada, Quebec-Windsor Corridor, Sweden, German ICE trains, France, Quasi-Direct Format (QDF) transport models, Armington Format trade models, Total transport demand models, Mode split models, Logit models, Price-Time Probit models, Dogit models, Inverse Power Transformation models, modal Induction, modal Diversion, Level-of-service (LOS) sensitivity, Response asymmetry, Cost damping, Cost amplification, Value of time (VOT), Rank Dependent Utility (RDU), Prospect theory, Attitude to outcome risk or uncertainty, Attitude to outcome, Attitude to distance, Autocorrelation, Heteroskedasticity, regression Coefficient sign change, Covariance among non orthogonal variables, statistical Causation or Correlation, Wiener-Granger causality approach, Beta-Lambda causality proposal, Forecasting, Long-distance passenger mode choice, Urban passenger mode choice, Freight mode choice, High Speed Rail market share and revenue forecasts, Distance response profiles.*

Journal of Economic Literature Classification: C-21, C-35, C-53, D-12, R-41.

0.2. Table of contents

<h1 style="text-align:center">List of Figures</h1>

<h1 style="text-align:center">List of Tables</h1>

0.3. Summary: the abstract with details added in italics

Setting a context for a discussion of rail demand function curvatures. As envisaged High Speed Rail (HSR) projects typically anticipate a reduction by half or more of existing rail Travel Time, ridership and revenue forecasts depend critically on the curvature of the rail demand curve defined with respect to both Fare and Time levels of service (LOS), in addition to Frequency. We discuss these curvature issues and derived forecasts within the Quasi-Direct Format (QDF) architecture in use since the US Northeast Corridor Transportation Project of the 1960's: a product of Total Market size and Mode split component models.

This project departed from the prevailing Direct Demand Format (DDF) practice of explaining T_m, the transport demand for any of M competing modes, by as many functions of all modal Fare and Time level of service (LOS) terms and of all intermediate and final activity (IFA) terms reflecting activities A_a or socio-economic Y_y determinants of travel.

The new QDF architecture core: (i) explained each modal demand T_m by a product of two functions, one determining total demand for all modes T_{TOT} [= $f_d(A_a, Y, U)$] and another the mode split p_m [= $U_m / \Sigma_m U_m$]; (ii) recognized two corresponding potential roles (on market size Induction and on modal Diversion) to all previously used variables, notably through the inclusion in the model for T_{TOT} of a coupling term U constructed from the denominator of the mode choice model; (iii) made the U_m attractiveness (Utility) function of each mode depend only upon own Fare and Travel Time terms (i.e. on the F_m and TT_m from the diagonals of LOS matrices of variables, and never on off-diagonal terms), thereby imposing the fiat *of additive separability of modal utilities and revoking the former possibilities of complementarity among the modes; (iv) combined this own-LOS diagonal "slavery" to the pre-existing own-OD pair "slavery" of the DDF whereby any flow $T_{m,ij}$ from origin i to destination j could only be explained by LOS or IFA variables obtaining one or more of the same own Origin-Destination pair indices (i.e. only by ij indexed values and never by ik or jn indexed values).*

In addition to this compounding of the IIA consistency built on spatial indices of all LOS and IFA terms with a new IIA consistency built on the LOS variables, the QDF structure maintained fixed forms for all variables involved in determining $T_m = [f_d(A_a, Y, U)] \bullet [U_m / \Sigma_m U_m]$. At first, the product involved two models of multiplicative form, an (unconstrained) Gravity-type specification for the Total trip component and a multiplicative Mode split model for the second component. This changed around 1975 when the Logit, emerging in Linear form despite earlier examples to the contrary, became the Mode split model of choice and the reference workhorse of a vast consulting industry in both transport and other fields.

The Logit component in a QDF structure admitting power transformations of variables. We study curvatures with Box-Cox transformations (BCT) applied to the variables of both components but our discussion focuses on the Logit piece used in representative QDF structures because modal diversion effects of new HSR services tend to dominate the slower long-term induction effects of increased market size, so that net discounted project results are driven there by Linear Logit specifications prevailing to this day, as they still do outside of transport applications.

Our strong focus on Box-Cox Logit form flexibility, effected within a QDF structure that is still consistent with the double IIA straightjackets just mentioned, aims at limiting paper length and is based on the fact that the demonstrated existence of the non-constant nature of the marginal utility linked to Fare and Travel Time terms (in such QDF structures respecting original IIA restrictions for mode split and generation-distribution components) is essentially unaffected by

generalizations of that structure that allow for potential complementarity among modes (rejecting additive separability of the utility of close travel alternatives) or among OD flows (rejecting the additive separability or independence among spatial transport or trade markets considered). We therefore refer only in passing, and merely for the sake of completeness, to these generalizations, such as the Generalized Box-Cox Logit (a workable form of the Universal or Mother Logit) and OD flow models admitting of spatial correlation among residuals: the additional role of cross-terms is specifically addressed in a previous paper on air demand.

The establishment of BCT non linearity and its use in forecasts. Our analysis starts with a reference summary of results from three Canadian mode choice models formulated in 1976-1978 and in 1992-1994 for the purposes of forecasting the effects of major infrastructure changes, respectively new airports in Southern Ontario and faster rail in the Quebec-Windsor Corridor. We emphasize how HSR revenue maximization of rail Fares and Speeds obtained under hypothesized Linear forms then yielded higher revenues than under data determined optimal Box-Cox forms even if their market shares were higher on relatively long distances. The rest of the paper interprets or positions these results among others and further documents Corridor model forecasts.

We first analyse, and find somewhat gratuitous, various popular objections raised to the strength of the accumulated evidence on the existence of non linearity in Logit models generally, notably (i) the presence of (observable) non spherical distributions of residuals due either to serial or directed autocorrelation, including spatial, or to heteroskedasticity; (ii) the existence of market segments (documented here with a Swedish case) and more generally of (unobservable) complete distributions of taste heterogeneity coefficients in Mixed Logit extensions (as demonstrated by recent Monte Carlo work on such specifications). We argue that form parameters are jointly determined with other system parameters and that, if some of the latter have distributions, the former should logically have them as well.

The *prima facie* meaning of gross BCT profiles in 40 models. Concerning interpretation, we point out that, although basic consumer demand theory does not constrain admissible values of BCT in Total demand and Logit Mode choice components of modal demand models, actual estimates are in fact generally compatible with "Cost damping" claims: (a) Time and Cost sensitivities (expressed as first partial derivatives of the demand function) typically fall with Distance in passenger and freight markets, except in urban passenger markets where Time sensitivity almost always increases; (ii) relative sensitivities (the Value of Time) always increase with Distance, irrespective of whether the absolute value of slopes falls at a decreasing (damped) or at an increasing (amplified) rate, in the 40-some surveyed models built by some 30 researchers for 10 countries. Such empirical regularities establish a fundamental difference between intercity and urban markets based on the second derivatives of demand functions.

*After an analysis of such models where the presence of flexible Box-Cox non linearity was detected for Time or Cost, we conclude that the validity of the cost damping claims is well served by the existence of flexible BCT non linearity and makes microeconomic sense. Making sense, here solely on basis of the Total market size and Mode choice determination models examined, simply means recognizing in cost damping (or amplification) claims the diminishing marginal utility of going further and the presence of curvature in utility maps governing rates of substitution among money and time characteristics of transport modes. In particular, it should not come as a surprise that **linearity of Logit models** is inconsistent with United Kingdom Department for Transport (UK Dft) cost damping concerns, as is apparently also the use of the **log-sum** aggregator in our form flexible QDF framework, its fully discrete variants included by implication.*

Gross and net power values. We further suggest that such real gross BCT power value sensitivity profiles estimated without taking the attitude to risk into account could in fact reflect two effects that, as shown in recent seminal work on Rank Dependent Utility (RDU), can be identified by products of power functions of Fare or Time, to wit a simple power to determine *the attitude to outcome risk (probability)* and a BCT power to determine the *attitude to outcome proper*. We also reinterpret recent models making successful use of interactions between a Distance variable raised to a simple power and a LOS variable (especially Travel Time) sometimes raised to a BCT power as identifying, also by a product of functions, an "attitude to Distance" that allows for an explicit breakdown of gross attitudinal parameters between *attitude to distance* and *attitude to outcome* elements.

We emphasise the existence of other benefits from the use of endogenous Box-Cox curvature determination, notably: (i) epistemological benefits from the establishment of statistical correlations whose existence, sign and size are not conditional on a priori mathematical form assumptions but are determined by the data, jointly with the existence, sign and size of form power parameters themselves; (ii) benefits from the knowledge that forecasts obtained from model parameters based on unconditional optimal form estimates differ from those obtained from non optimal conditional form parameter estimates for otherwise similar specifications, as we demonstrate by both analytical and simulation methods; (iii) benefits from computations of consumer surplus under correct curvatures.

UK Dft concerns are therefore valid and matter: linear and non linear demand model forms imply more accurate and quite different HSR market share profiles over long distances; and profiles derived from demonstrably non linear models are of critical relevance to HSR revenue forecasts because they often *imply relatively and absolutely higher rail passenger flows over long distances but lower project revenues.*

Robustness of results from estimated curvatures. We imply overall that, beyond mere fit and other demonstrated benefits, untested linear forms of Standard Logit utility function variables are theoretically unexpected as representations of price-time utility maps, statistically unsustainable in many samples ranges where *prima facie* gross cost or time damping or amplification prevail in absolute and relative senses, practically biased as conditional bargaining games to play with data, and often demonstrably unsound or misleading in the production of HSR passenger and revenue forecasts, and no doubt elsewhere as well.

1. Implications of non linearity of Logit utility functions[1]

Curvature and forecasting: the importance of non linearity. As HSR investments basically divide rail travel time by half, the key to correct forecasts of the number of increased passengers and revenues obviously lies in the curvature of the demand curve fitted here primarily, but not exclusively, with Box-Cox (1964) transformations (BCT). In practice, as most of the short term effects of HSR happen through modal diversion, rather than through induced growth in the market ─ of more limited import because its discounted present value is typically relatively small ─ that key lies in the Mode choice model, typically a Multinomial Logit (MNL) model or a close substitute. However, the curvature issue would be the same with any model explaining choices by reference to the Utility of alternatives defined by modal characteristics such as Time, Cost and Frequency of service and raising the question of the sensitivity of the market to changes in those Level-of-Service (LOS) factors, among others.

To probe the issue of the existence, strength and relevance of non linearity of LOS factors, we adopt a modeling architecture called the Quasi-Direct Framework (QDF) which admits of many possible cores and select the suitable Logit core for the main body of our probe, but without neglecting in a second step information extracted from enriched Dogit and Inverse Power Transformation-Logit core examples. Results from more than 40 models provide a context to position results from three models designed specifically to study HSR potential in Canada.

The establishment of non linearity in the Canadian and Swedish markets. We first recall results from the first one, published in 1978, written with a view to forecasting the impact of a new airport in Southern Ontario in an environment where passenger rail potential in the Quebec City-Windsor Corridor of Canada was already of explicit concern (Wills *et al.*, 1976; Rea *et al.*, 1977).

We then report on two unpublished papers written with databases collected specifically to study various HSR options in that Corridor of 1150 km by 100 km, a band of relatively high population density where, counting intervening cities such as Montreal, Ottawa, Toronto and Hamilton, there are as many inhabitants (up to 16,0 million, or 50% of the Canadian population) as in The Netherlands, an area ($42\,000$ km^2) half of the Corridor size. The first of the latter papers was produced as part of a federal-provincial official and public inquiry made around 1990 and the second as part of a private study conducted by Air Canada and Canadian Pacific Rail (CP Rail) in 1994.

We complement these Canadian findings with those of a fourth paper, also unpublished, probing a Swedish national model and focused on the distinction between market segmentation and non linearity. Coverage of Swedish work opens the door to a discussion of two particular objections to the existence of non linearity: the presence of market population segments ─ and even of complete distributions of individual taste differences ─ and the presence of non spherical residuals.

In these four models, the analysis of the implications of non linearity is established in the usual way by a comparison of overall measures of fit and of elasticities obtained with different forms. Best fit forms are shown to be typically non linear and to imply very different elasticities from linear ones, at least for Time and Cost factors, but without neglecting in passing the non linearity of Frequency of service.

But, in addition to providing such indicators, the authors of the Air Canada-CP Rail study also formally optimized train speeds and fares in their model and, interestingly, found that HSR investments forecasts based on (optimal) non linear forms were less profitable than if they were based on (rejected) linear utility functions. This simulation analysis was made observation by observation, but the detailed results were not shown on graphs. They were implicitly related to Distance in aggregated constructs for a few important origin-destination (OD) pairs, or markets, but the individually calculated forecast differences between models differing only in form were not further graphed against Distance, or any other Corridor factor, to provide a sense of the total picture.

Making sense of actual non linear form estimates. The difficult question of the values to be expected for BCT parameters in such studies is then raised as an attempt to answer "Cost damping" queries recently put in the United Kingdom. The existence and nature of Cost damping and of its opposite, Cost amplification, is pinpointed on the basis of our survey findings concerning BCT Time and Cost parameter estimates from models pertaining to intercity passenger or freight and to urban passenger Logit applications as well.

Overall, the survey results reveal two extraordinary regularities of the ranges of estimates: first, they show that damping predominates, *i.e.* that demand slopes with respect to Time and Cost typically fall with distance at a decreasing rate both absolutely and relatively; second, they also indicate that the amplification exceptions, *i.e.* slopes falling at an increasing rate with distance, do not occur at random but are found in almost all urban and rarely in some of the longest intercity markets.

We argue that LOS risk, as formalized in Rank Dependent Utility (RDU) approaches, might explain these cases of amplification, and notably the clear specificity of urban model results, as could also the growing practice of multiplication of LOS factors by simple powers of Distance interaction terms, a practice newly interpreted here as the measurement of an attitude to Distance to be distinguished from the reaction to Distance already accounted for in the levels of the LOS variables.

Revenue forecasting implications of non linearity. We argue that, to understand implications of a given model, there is in practice no alternative to the detailed simulation approach and to an analysis of forecast differences by relating them to a variable of interest, such as Distance. We therefore apply the new label "Distance profile" to analyses formerly made in two European markets and proceed to perform similar "Distance profile" analyses of Canadian results obtained from the official government-approved Corridor model.

Overall evidence from such simulations for business trips in the Corridor, made obvious by Distance profiles, is that optimal non linear forms rarely produce higher passenger mode share gains than presumed linear forms. And when they do for relatively long trips, these may occur simultaneously with lower ones for relatively short trips where most potential clients are found, thereby implying lower overall expected HSR revenue gains, a situation we call linear model over-prediction bias.

But proper forecasts are not the only benefit derived from tests of functional form: knowledge that results are derived from "form-sensitive" correlations also matter for the Quebec-Windsor Corridor project in an environment where linear Logit models occupy the field, in transport and elsewhere.

2. The QDF framework with a Logit core

Demand framework or format. Mode choice, despite its dominance in project assessment, is only the principal dimension of Demand determination. It needs to be considered within a complete representative demand framework that matches current best demand modelling practice.

In the early 1960's, transport demand formulations at first simply enriched existing microeconomic consumer demand system formulations where Income and Prices of all goods as a rule appeared in the demand function for each good. The enrichments recognized that the derived demand for transport should depend on levels of intermediate and final Activities, and not just on Income, and that Service conditions of all M modes mattered as much as their Prices. This "Direct Demand" (DDF) format, may be written for any origin-destination (OD) pair (o = 1, … , O; d = 1, … , D):

(0-A) $$T_{od,m} = f(A_o, A_d; U_{od,1}, ..., U_{od,M}), \qquad m = 1, ... , M;$$

with

(0-B) $$U_{od,m} = f\left[(Cost_{od,m}, Service_{od,m}); X_{Socio-economic}\right],$$

and we remain for the moment intentionally vague concerning the units of measurement of money and time characteristics of the modes.

As such DDF formulations and their variants, typically of multiplicative form, posed considerable problems of collinearity and often yielded "unreasonable" signs for most coefficients of modal characteristics, a complete family of new models arose, in the context of and within the Northeast Corridor Transportation Project (NECTP), which progressively resolved many problems satisfactorily by replacing Direct framework (0) by Product framework (1) combining Total Demand and Mode Split components. The resulting "Quasi-Direct" format (QDF) design basically retained the market scale determination component of DDF models and recombined modal Prices and Service levels within a Mode split component that eventually proved amenable to both aggregate and discrete formulations. In addition, the components could be coupled.

During the transition period from Direct demand by mode to Quasi-direct demand by mode, many formulations of the modal choice component were tried[2], to wit: ratios of Levels of service (LOS) variables, such as Cost and Time characteristics of each mode *m* to "best mode" levels; "abstract mode" generic constraints imposed on LOS coefficients with a view to forecasting the potential demand for supposedly "new" modes[3]. Most formulations maintained the multiplicative form of the equations[4].

As a consequence, direct demand models became a minority stream either in simple multiplicative form or in more sophisticated garbs. In the former, they are still adopted in local studies focussed on the demand for a single mode and modified heroically in *ad hoc* fashion, for instance to make the elasticities variable or sensitive to "competition" (cross service terms), as in Wardman (1997). In the latter garbs, they have borrowed the theoretical sophistication of the "trans-log" (Christensen *at al.*, 1971), a second order approximation adding to the simple product of variables all of their

---

[2] Different new approaches were gathered by Quandt (1970) in an excellent book.

[3] The reader might consult Lave (1972) for a typical discussion reflecting this confusion to be clarified later when formulations based on BCT made it possible to nest most of the individual specifications.

[4] In Appendix A of CRA (1972), unfortunately never published, authors tried to link the form of the aggregate demand equations to the structure of underlying Quadratic, Stone-Geary and Log-Linear utility functions, but their derivation was not successful: "Even for log-linear utility functions, the fact that the commodity "transportation" is logically expressed in terms of a number of attributes leads to an intractable specification of the demand function" (p. A-4).

interactions (*e.g.* Oum & Gillen, 1983), a problematic undertaking if all qualitative characteristics of the modes should be multiplied in as well; and the issue will not go away if a Linear Expenditure System formulation is adapted to all these service terms (Andrikopoulos & Brox, 1990).

But the rise of mode choice models proper, especially of the discrete kind, soon occupied the largest part of the research space and almost all of the practical project forecasting space, as it still does to-day. Some discrete choice specifications were soon extended to the explanation of the amount of trip making but we will not deal with them explicitly here.

We discuss non linearity within the QDF framework, whereby T_{odm}, the Demand for a particular mode *m* from *i* to *j*, is obtained as a product of a model of the Total transport market by all modes T_{odTOT} and a model of Modal split P_{odm}. Neglecting *od* subscripts, this may be expressed as:

(1) $$T_m = \{T_{TOT}\} \bullet \{P_m\},$$

or, more explicitly, as:

(2) $$T_m = \left\{ f(A_c, A_d, U) \right\} \bullet \left\{ U_m / \sum_m U_m \right\}, \quad m = 1, ..., M,$$

where the model of Total demand by all modes contains vectors of activity variables A_c and A_d, such as Population and Income, and an index U of the utility of travel often called the "inclusive" value of the modes is automatically used as coupling term. It is simply the denominator of the mode split model where each Modal utility U_m term summarizes the attractiveness of a particular mode:

(3) $$U = \sum_m U_m, \; U_m \geq 0$$

This framework admits of many Modal utility "cores" characterizing both Split and Coupling where the Split may come from either aggregate (explaining market shares) or discrete (explaining categorical individual choices) applications, as shown in Tran & Gaudry (2008g).

Logit quantities or cores. In the Multinomial Logit[5] core, the split model *p(i)* explaining the market share or choice probability of the ith of *M* modes, is:

(4) $$p(i) = \frac{\exp(V_i)}{\sum_{j \in C} \exp(V_j)} \qquad , i, j = 1, ..., M$$

where the V_i functions are the so-called "representative utility functions" (RUF) of the modes and consist in modal characteristics, such as Time, Cost and Frequency of service that we focus on here, and in socio-economic traits of travellers. We are mostly interested in Logit core quantities $\exp(V_i)$ based on the MNL but Section 7 defines enriched alternate cores. And any core might rely on one or more RUF specifications.

Representative utility functions. For the RUF, we first adopt the classical tradition whereby representative utility functions are specified with respect only to own-mode network variables X^i_n and to socio-economic variable X_s common to all alternatives:

(5) $$V_i = f(X^i_{Network}, X_{Socio-economic}).$$

We will relax later this constrained view based on the stunning assumption that the attractiveness of a mode does not depend on the characteristics of other modes and is therefore said to obtain

---

[5] As pointed out by Maddala (1983, p. 42), there is no algebraic difference between the Multinomial Logit (MNL) and the "Conditional" Logit proposed by McFadden (1974), which just happened to make choices depend on LOS variables in addition to Socio-economic variables in an urban mode choice problem. We adopt the MNL term.

"separable" utility. But inclusion in (5) of network characteristics of other modes than the ith, an enrichment possibility touched on below in passing and detailed in Appendix B, is not central to our form concerns and would not substantially modify our central conclusions on form matters, as demonstrated at length elsewhere in the related context of the modelling of competition among airports (Gaudry, 2010).

Model components with fixed and stochastic parts. To discuss non linearity in the determination of the Demand for a mode T_m, we first need to specify the fixed and stochastic parts of both building blocks T_{TOT} and V_i. We do this with contrasted Box-Cox Classical and Box-Cox Logit regression specifications, respectively stated for each model component as:

(6-A) $\qquad T_{TOT,t}^{(\lambda_{TOT})} = \beta_0 + \sum_k \beta_k X_{k,t}^{(\lambda_k)} + u_t,$ $\qquad$ where t is an observation subscript and

u_t is IID Normally with variance $\sigma^2[f(Z)_t]$;
and each set Z of variables $Z_1, \ldots, Z_H$ may include some X_k variables and $[f(Z)_t] \geq 1$;

and

(6-B) $\qquad V_{in} = \beta_{i0} + \sum_k \beta_{ik} X_{ikn}^{(\lambda_k)} + e_{in},$ $\qquad$ where n is an observation subscript and

the e_{in} of all M modes are IID Weibull, each with variance $\pi^2\mu_i^2[f_i(Z_i)_n]/6$ and the μ_i stands for the scale factor of the Weibull distribution $\{f(u_i) = (1/\mu_i)[\exp(-u_i/\mu_i)][\exp(-\exp(-u_i/\mu_i))]\}$;
and each set Z_i of variables $Z_{i1}, \ldots, Z_{iH}$ may include some X_{ik} variables and all $[f_i(Z_i)_n] \geq 1$;

where we use W_f, the commonly named "Box-Cox Transformation" (BCT) applicable to a variable $\dot{W}_f$ or to a function $W_f^\bullet$. The BCT, the best known and most used of the power transformations, was formulated in direct form by Box & Cox (1964) and introduced in inverse form by Gaudry (1981):

Direct Power Transformations (DPT)		Inverse Power Transformations (IPT)	
$W_f\left(\lambda_f, \mu_f\right)$	Conditions[6]	$W_f\left(\lambda_f, \mu_f\right)^{-1}$	Conditions[7]
BC		BCG[8]	
$\dfrac{W_f^{\lambda_f} - 1}{\lambda_f}$	$\lambda_f \neq 0, W_f > 0$	$\left(\lambda_f W_f + 1\right)^{1/\lambda_f}$	$\lambda_f \neq 0, \left(W_f + 1\right) > 0$
$\ln W_f$	$\lambda_f \to 0, W_f > 0$	$\exp W_f$	$\lambda_f \to 0, W_f > 0$
BT		BTG	
$\dfrac{\left(W_f + \mu_f\right)^{\lambda_f} - 1}{\lambda_f}$	$\lambda_f \neq 0, \left(W_f + \mu_f\right) > 0$	$\left(\lambda_f W_f + 1\right)^{1/\lambda_f} - \mu_f$	$\lambda_f = 0, \left(\lambda_f W_f + 1\right) > 0$
$\ln\left(W_f + \mu_f\right)$	$\lambda_f \to 0, \left(W_f + \mu_f\right) > 0$	$\exp\left(W_f\right) - \mu_f$	$\lambda_f \to 0, W_f > 0$

(6-C) refers to the table above.

At this point, we need to be more explicit concerning the treatment of error variance discussed at length further below: we provisionally assume constancy of error variance, called

---

[6] Under certain conditions, it is also possible to transform an explanatory variable that contains some zero observations but is not a Boolean (dummy) variable.
[7] This development of inverses in early 1978 was partly aimed at transforming the dependent variables of probability and share models in the same way as one transforms the dependent variable in (6-A), a predicament if probabilities must directly sum to one. But this constraint can be indirectly met if the transformations are applied to right hand side quantities, Modal utility functions for instance. They bring out in probability models the possibility that, in addition to belonging to LOS variables, curvature also belongs to each Modal utility term, an issue discussed further in Section 7.
[8] The label was intentionally invented by students as a pun on the French name of a vaccine given to children.

homoskedasticity[9], despite the fact that both (6-A) and (6-B) as specified have very general built-in heteroskedasticity formulations.

In the former, the assumption of homoskedasticity $\{\sigma^2\}$ requires that $\{[f(Z)_t]=1\}$ or, alternatively, dividing all variables (dependent and independent including the intercept) by $\{[f(Z)_t]^{1/2}\}$, an expression proposed in Gaudry & Dagenais (1979b) which conveniently includes many classical fixed-form heteroskedasticity specifications as nested special cases obtained by specializing values of the BCT and of the δ_m coefficients:

$$(6\text{-D}) \qquad f(Z)_t = \exp\left[\sum_m \delta_m Z_{mt}^{(\lambda_{Z_m})}\right], \text{ where } f(Z_t) \text{ may contain one } Z_{mt} = X_{kt} \quad .$$

In the latter, where we have used an analogous formulation for each representative Modal utility, homoskedasticity requires that the $\{\mu_i\}$ and all $\{[f_i(Z_i)_n]\}$ equal 1 or, alternatively, demands dividing all variables including modal constants of utility functions by $\{[\mu_i[f_i(Z_i)_n]]^{1/2}\}$ where

$$(6\text{-E}) \qquad f(Z_i)_t = \exp\left[\sum_m \delta_{i,m} Z_{i,mt}^{(\lambda_{Z_m})}\right], \text{ and } f(Z_{i,t}) \text{ may contain one } Z_{i,mt} = X_{i,kt} \quad .$$

The reasoning behind (6-E) is simply that, if an heterogeneous error variance equal to $\{\pi^2\mu_i^2/6\}$ requires dividing all regressors of the Modal utility by $\{\mu_i\}$ to obtain homoskedasticity $\{\pi^2/6\}$, an error variance equal to $\{\pi^2\mu_i^2[f_i(Z_i)_n]/6\}$ likewise requires[10] dividing instead by $\{\mu_i[f_i(Z_i)_n]^{1/2}\}$ to obtain homogeneity $\{\pi^2/6\}$ of the variance of each Weibull distributed error, as will be done below.

Transformations, model components and form structure. QDF formats allocate LOS variables jointly to Modal utilities U_m and to the coupling term U. So changes in LOS impact Modal demand T_m through both paths. In these circumstances, the study of curvatures of demand functions with respect to Time and Fare LOS and U variables transformed by BCT is directly dependent on the form structure $(\lambda_U, \lambda_T, \lambda_F)$ assumed or estimated.

As many practitioners associate the *log-sum* term, obtained if $(\lambda_U = 0)$, and linearity of the other terms $(\lambda_T = \lambda_F = 1)$ even if these need not be joined, it is useful to define a reference case triplet in this way: we will see that it corresponds to demand slopes that are completely independent from the LOS variables and posses no curvature.

Our *ceteris paribus* analyses of effects of changes in LOS variables do not mean that the application of BCT within a model is solely limited to form structure elements $(\lambda_U, \lambda_T, \lambda_F)$.

The QDF structure with a Logit core and some known models from the 1960's. For our ends, there would be no conceptual gains in replacing the aggregate model of the Total demand for all modes T_{TOT} in the QDF structure (1), a proper demand function, by a probabilistic component. It is the coupling that makes for a proper demand function because Modal utilities determine both Split among competing modes and the overall attractiveness of available alternatives sustaining the Total demand for travel. Naturally, other structures also explaining total demand, and consequently reaching beyond the mere explanation of splits, can be built and constitute proper demand functions where Total demand is endogenous.

---

[9] Note that we reject in English the spelling "*homoscedastic*", as recommended by McCullough (1985).

[10] This may be understood as a pre-specification of "corrected" representative utilities with independent *homoskedastic* Weibull errors discussed at length in Palma & Thisse (1987).

An advantage of the QDF framework is that it nests many intercity models formulated for the NECTP and allows for a number of Logit core variants as well. For instance, grouping (6-A)-(6-B) and (6-C), the framework may be fleshed out, neglecting error terms, as:

$$(7\text{-A}) \quad T_m = \left\{ \beta_0 + \sum_{k=1}^{K} \beta_k X_k^{(\lambda_{X_k})} + \beta_U \left[\sum_{j=1}^{M} e^{\beta_{j0} + \sum_n \beta_{jn} X_{jn}^{(\lambda_{jn})} + \sum_s \beta_{js} X_{js}^{(\lambda_{js})}} \right]^{(\lambda_U)} \right\}^{(\lambda_{TOT})^{-1}} \cdot \left\{ \frac{e^{\beta_{m0} + \sum_n \beta_{mn} X_{mn}^{(\lambda_{mn})} + \sum_s \beta_{ms} X_{ms}^{(\lambda_{ms})}}}{\sum_{j=1}^{M} e^{\beta_{j0} + \sum_n \beta_{jn} X_{jn}^{(\lambda_{jn})} + \sum_s \beta_{js} X_{js}^{(\lambda_{js})}}} \right\}$$

which, if all BCT are set to zero, yields the seminal multiplicative model by McLynn *et al.* (1968):

$$(7\text{-B}) \quad T_m = \left\{ \prod_{k=1}^{K} X_k^{\beta_k} \right\} \cdot \left\{ \sum_{j=1}^{M} \left[\prod_n X_n^{\beta_{jn}} \prod_s X_s^{\beta_{js}} \right] \right\}^{\beta_U} \cdot \left\{ \frac{\left[\prod_n X_n^{\beta_{mn}} \prod_s X_s^{\beta_{ms}} \right]}{\sum_{j=1}^{M} \left[\prod_n X_n^{\beta_{jn}} \prod_s X_s^{\beta_{js}} \right]} \right\}$$

and closely related formulations[11] by McLynn & Woronka (1969). Such multiplicative formulations resemble, but should not be confused with, another multiplicative design developed at the same period as the QDF by Armington (1969) and favoured since by trade flow analysts[12], but that was only relatively recently applied to transport flows (*e.g.* Gillen *et al.*, 1999, 2001).

A third format. In that Armington demand format, the U_m functions are raised to a common simple power, σ for instance

$$(6\text{-F}) \qquad\qquad U_m = \left[Cost_m \right]^{-\sigma}$$

for a case with a single modal characteristic, where this power σ is the Constant Elasticity of Substitution (CES) across alternatives specific to the total demand market considered and the coupling term (3), called "Armington aggregator", now obtains a different composition, namely

$$(6\text{-G}) \qquad\qquad U \equiv \left\{ \sum_{m=1}^{M} \left[\delta_m \cdot Cost_m^{-\sigma} \right] \right\}^{-1/\sigma}$$

where the δ_m modal weights have to be determined somehow[13].

DDF, QDF and Armington formats are in competition but, fortunately, Anderson & Palma (2000) and Palma & Sanchez (1998) have provided a theoretical specification within which the last two are nested and should eventually be compared. They have shown that, if one uses the following BCT specification $\lambda = (1 - \eta)$ on (5), the systematic parts of the V_m functions found in (6-B):

$$(6\text{-H}) \qquad\qquad f\left(V_i / \mu \right) = \frac{1 - \left[V_i / \mu \right]^{1-\eta}}{1 - \eta}, \quad 0 \le \eta \le 1$$

where the μ_i from (6-B) are constrained equal, then setting $\eta = 0$ yields $[\exp(V_i)]$, the Logit Quantity (4), and $\eta = 1$ yields $[V_i]^{-(1+\mu)/\mu} \equiv [V_i]^{-\sigma}$, the CES Quantity (6-F).

---

[11] For a detailed discussion of many such cases nested into the QDF and of the related mode-abstract model by Quandt & Baumol (1966) and its successors (Young, 1969), the reader may consult Gaudry & Wills (1978).

[12] That literature has to specify structures in which each commodity (like a modal flow) is traded everywhere and in both directions (in spite of the theory of comparative advantage) in quantities that depend on all individual commodity Prices (or, in transport, eventually on their Service quality-adjusted generalization).

[13] Gillen *et al.* (1999, 2001) assume that they are equal across alternatives but sample market shares are sometimes preferred with similar arbitrariness.

14

Unfortunately, the CES specification (6-F)-(6-G), originally based solely on Cost, is hard to extend to Service quality variables, except perhaps by *ad hoc* rigidly multiplicative quality adjustments to Cost[14]. Moreover, even without the issue of the unavoidable incorporation of all LOS factors, the estimation of simple power functions poses problems of its own documented in Appendix A. Their estimation (i) may lead to degenerate solutions, (ii) does not preserve the ordering of the data and (iii) is not continuous at zero, a critical area for model values, nesting and taxonomy.

We surmise that those unavoidable LOS specification issues and the difficulty of obtaining an endogenous determination of weights in (6-G), combined with the supplementary care required to estimate simple power functions, will keep the Armington format a rare occurrence unless Box-Cox tests based on (6-H), absent even from Anderson & Palma (2000), show that it should prevail.

Sequential estimation of QDF component models. For the well tried QDF with the Logit core design (7-A) adopted here, we assume without loss of generality that the Multinomial Logit Mode choice component is estimated first and the Total demand component second. This sequential approach matches prevailing disjoint practice and conveniently allows refined focus on the issues pertaining to each component, such as form and stochastic specification. But there might be a hidden cost beyond that of efficiency loss to the extent that, over time since 1975, socioeconomic characteristics have migrated to the mode choice model without demonstration of their relative merits in each component. Joint estimation would favour proper assignment of such variables to either or both components.

Some authors of course do prefer to gain efficiency by the direct estimation of the Log likelihood of T_m defined by (2), although this can make for a very complicated maximization problem. For instance, Laferrière (1988, 1999) estimated the Air Path demand resulting from the product of a Total Air market and of an Air Path choice model for the domestic Canadian air market but took many of the mode choice form parameters considered, based on IPT-Logit form (12-A) below[15], as given.

In any case, the product of Maximum likelihood estimates for each component of (7-A) yields Maximum likelihood estimates for derived results calculated from the product of chosen estimated pairs of models, as effected by algorithms[16] that allow such QDF combinations based not only on Logit cores but on many other cores as well[17].

But a more important practical limitation of our chosen (7-A) should be mentioned, that of its "double IIA consistency", which is not a necessary feature of all constructs of type (2)-(3).

Doubled-up IIA blinkers, or reinforced diagonalism. We have pointed out in the third[18] section of a previous paper on competition among airports (Gaudry, 2010) from which some of our current tables are drawn or inspired, that the QDF structure (7-A) is in fact consistent with the

---

[14] Liked by theoreticians (*e.g.* Tirole, 1988, Ch. 2) but which will suffer from lack of flexibility when aligned against a flexible form Box-Cox Logit where the optimal form is generally no more logarithmic than it is linear, as we document extensively below.

[15] As documented in Section 7, the Inverse Power Transformation-Logit core makes it possible to identify specific constants for all alternatives (here air paths), in contrast with Logit cores where the identification of all intercepts important to the author is impossible.

[16] The QDF algorithm (Tran & Gaudry, 2008g, 2010) allows for 20 such possible products. See Table 12 below.

[17] Section 7 illustrates two Dogit (Gaudry & Dagenais, 1979a) and two Inverse Power Transformation-Logit (Gaudry, 1981) cores.

[18] The first section develops input-output matrix representation of transport flows; the second addresses conceptual structure issues in the determination of LOS performance in networks; the third, summarized in Appendix D, emphasizes the importance of moving away from OD-type consistency with IIA to understand transport or trade flows.

Independence for Irrelevant Alternatives (IIA) axiom of choice theory (Luce, 1959) in two senses: first, with a Standard Box-Cox Logit utility function based on (5), only *own LOS* variables appear in each Modal utility function; second, in the complete structure, only *own-OD flow* indexed variables are admitted.

The focus of that third section is on breaking up this "diagonal slavery" either with a Generalized Box-Cox utility function specification of type (8-A) presented as a workable form of the Universal or Mother Logit (McFadden, 1975) in the Mode choice piece, or by taking spatial autocorrelation into account in the Total demand piece: both methods introduce cross-diagonal terms that move away from IIA diagonalism in one or both senses of this expression.

A key finding is that, in all known examples, introducing non linearity to LOS variables in Logit models causes higher log-likelihood gains than further adding suitably non linear cross-diagonal terms[19], a result that is without surprise if the vast literature on systems of demand equations, where cross-diagonal effects are rarely if ever dominant, is recalled. Other key results on the rejection of separability among competing OD flows, so important to tourism and air markets, are not as directly relevant here but are also consistent with the view that getting the right form on the diagonal matrices of LOS terms is a more important practical matter than further breaking IIA consistency with the addition of off-diagonal terms, be they themselves of the proper data-determined form or not.

In a demand system, liberation from diagonal slavery is less important than liberation from incorrect form.

---

[19] For instance, in the freight model used for Figure 11 below, and coming from Gaudry *et. al.* (1998), the introduction on two BCT on Distance and own Price increases the log-likelihood by 46 points (for 2 degrees of freedom) and the further addition of two cross Prices by 25 points (for 4 degrees of freedom), but the BCT on own price changes only from -1,83 to -1.89. This Generalized Box-Cox form, where the three modal prices appear in all Modal utility functions, is used by the French ministry for Transport since 2006.

3. The establishment of non linearity in Logit models

3.1. A starting point: is marginal utility constant, really?

The first question raised by classical Logit representative utility functions, in the past and now, is whether their linearity can be credible, as assumed long ago (*e.g.* Domencich & McFadden, 1968, 1975) in spite of previous work with logarithmic specifications based on "the fact that this type of [logarithmic] relationship has often proved successful in other types of demand analysis, and by the fact that the scatter diagrams appeared to support the hypothesis in this case" (Warner, 1962, p.27). Testing is warranted, if only because linearity is as unexpected in Nature as in microeconomic consumer demand theory.

On this, note that the more general Box-Cox specification (6-B) makes the effect of a network service improvement depend on the level of the modified characteristic, as shown in Table 1. This implies that the impact of a 10 minute change in travel time is not the same for a short and for a long trip; it also makes derived marginal rates of substitution between time and money (values of time or VOT) vary both across modes (due at least to different sample levels of the characteristics) and with the amount of time saved. It can therefore in principle avoid much combinatorial market segmentation used by energetic piecewise linear approximation analysts to obtain reasonable and variable trade-offs by class of distance, fare level, departure frequency, income, *etc.*

Table 1. Marginal utility in representative utility functions of a Standard Box-Cox Logit model

Form	$\dfrac{\partial V_m}{\partial X_{mk}} = \beta_{mk} X_{mk}^{(\lambda_{mk}-1)}$	$\dfrac{\partial V_m^2}{\partial X_{mk}^2} = \beta_{mk}\left(\lambda_{mk}-1\right) X_{mk}^{(\lambda_{mk}-2)}$	*Returns*
$\lambda = -1$	β_{mk}/X_{mk}^2	$-2\beta_{mk}/X_{mk}^3$	Decreasing
$\lambda = 0$	β_{mk}/X_{mk}	$-\beta_{mk}/X_{mk}^2$	Decreasing
$\lambda = 1$	β_{mk}	0	Constant
$\lambda = 2$	$\beta_{mk} X_{mk}$	β_{mk}	Increasing

It is also worth remembering that, as shown in Figure 1, non linearity implies that the market probability (or share) response curve to any change in modal characteristics is asymmetric with respect to its inflexion point and, contrary to the linear case, does not have the inflexion point at 0,5.

Figure 1. Classical Linear-Logit *vs* Standard Box-Cox-Logit Responses
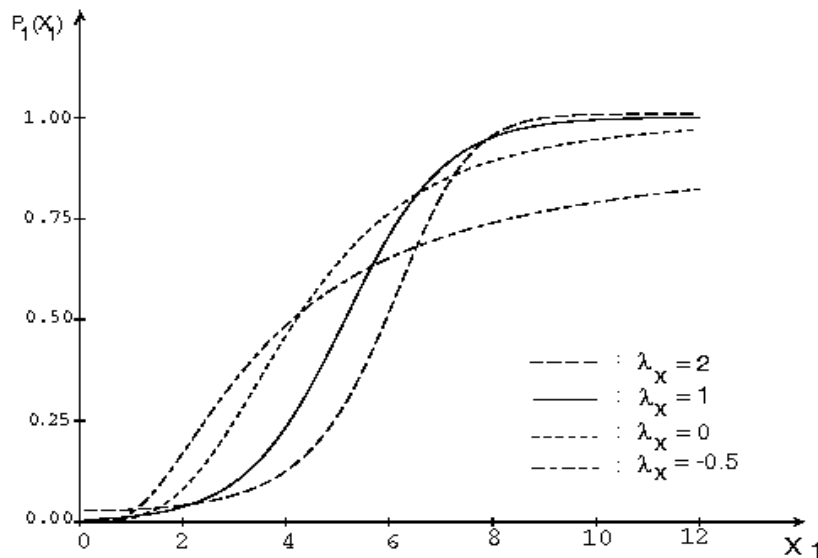
In Figure 1, the x-axis does not designate the representative utility function V_1 from (4), which always yields a symmetric response curve of the choice probability, but a particular variable X_1 included in V_1, in this case Air Time. So, if one believes that a response curve so defined for a mode, diffusion or learning process, is not symmetric with respect to a particular factor, this belief implies an underlying non linear utility function, and conversely. In practice, it might not really matter whether one believes that the function is linear or not because only tests should make a result acceptable: if one's belief is that the function is non linear, the belief should be put to the test of the data; and, if one has no idea of the true form, only tests can determine the nature of the relationship. The practical issue of curvature, as formulated by Warner, is clearly at least one of "scatter diagram" fit; but it also goes much deeper, as we shall later demonstrate.

But, beyond asymmetry of response curves, one can observe other interesting features in Figure 1 where each asymmetric curve implies, over the range of variable X_1 favourable to mode 1, a specific profile of differences between market shares (or probabilities) predicted by any two model variants differing only in form, for instance the reference linear and any maintained non linear models. Denoting by Δp_{X_1} those differences calculated with respect to a certain variable X_1, and giving to plots of Δp_{X_1} against X_1 the name "X_1 -Profiles of Δp_{X_1}", positive values within such profiles mean that the reference linear model under-predicts market shares and negative values that it over-predicts them.

In Figure 1, the three X_1 -Profiles of Δp_{X_1} derivable from the reference and the three maintained non linear response curves shown happen to generate two sequences of alternating sign ranges, the second more difficult to identify visually:

(i) under $\lambda_X > 1$ in the maintained model, the sequence: [under-prediction $(\Delta p_n > 0)$ at relatively low values of X_1; over-prediction $(\Delta p_n < 0)$ at mid-range; under-prediction $(\Delta p_n > 0)$ at relatively high values of X_1];

(ii) under $\lambda_X < 1$ in the maintained model, the sequence: [over-prediction $(\Delta p_n < 0)$ at relatively low values of X_1; under-prediction $(\Delta p_n > 0)$ at mid-range; over prediction $(\Delta p_n < 0)$ at relatively high values of X_1].

But it is important to realize that each X_1 -Profile of Δp_{X_1} is unique and that the sequences generated from Figure 1 response curves are specific to the model at hand and do not depend systematically on the symmetry pivot $\lambda_X = 1$, as will be demonstrated further on.

For Figure 1, the model at hand, estimated only for the purposes of the figure, was a simplistic Air vs Car binomial aggregate application with utility functions $V_1 = \beta_0 + \beta_1 X_1^{(\lambda_1)}$ and $V_2 = 0$, with parameters re-estimated for different values of λ_1 with a sample of 120 intercity OD pairs for Canada in 1976 described in Gaudry (1990 or 1993); and with X_1 = (Time by car/Time by plane). It is an artificial model and only the 120 OD pairs for which observations on the 4 intercity modes are available are chosen. The full sample contains 40 more OD pairs for which one of the 3 modes other than car is missing because the mode is unavailable.

The decision to neglect these 40 observations was made because it is notoriously difficult, but possible, to estimate models of shares in which some modes are sometimes unavailable (as long as one of them is common) and even more difficult to estimate share models where null observations on modal shares are kept in the sample: a formulation (Dagenais, 1986) requiring trivariate normal integrals in a 3-mode specification remains to-day as unpromising as it was in 1986, except perhaps to Probit fans.

3.2. A first aggregate model for Canada (1978)

The question of nonlinearity in Logit models was originally asked in an intercity model designed primarily to forecast the impact of the location of a new airport in Southern Ontario and to improve existing rail demand forecasts based on PERAM (Rea *et al*., 1977; Transport Canada, 1979). The data base pertained to the four intercity modes linking 92 Canadian OD pairs for the year 1972.

The mode choice model was defined with generic specifications for the regression coefficients of the three service variables: Fare, Travel time and Departure frequency. Numerous formulations of these variables, including Box-Tukey origin shifts, were tested. Results in Table 2 show dramatic log-likelihood gains of Box-Cox forms over fixed linear or logarithmic forms.

Table 2. Four-mode Box-Cox Logit equation results (Canada, 92 city-pairs, 1972)

Gaudry & Wills (1978)	*Column* *Case* *Original reference*	*1* *Linear* *F*	*2* *Log* *E*	*3* *Box-Cox* *D*	*4* *Box-Cox* *C*
Fare (per O-D)					
	Coefficient	-0,0008	-2,9653	-2,2254	-1,8274
	Conditional *t*-statistic	(-11,31)	(-15,57)	(-18,70)	(-17,09)
Travel time (per OD)					
	Coefficient	-0,0141	-1,3148	-0,3605	-0,8358
	Conditional *t*-statistic	(-9,09)	(-5,57)	(-4,14)	(-4,59)
Frequency of service (per OD, except for car)					
	Coefficient	0,0114	0,4221	0,1331	13,50
	Conditional *t*-statistic	(3,52)	(4,60)	(5,02)	(5,36)
Box-Cox λ on Fare		**1,00**	**0,00**	**-0,1930**	**-0,2626**
Box-Cox λ on Travel time		**1,00**	**0,00**	**-0,1930**	**-0,0513**
Box-Cox λ on Frequency		**1,00**	**0,00**	**-0,1930**	**+0,5712**
Log-likelihood		456,32	528,71	532,97	538,66
Degrees of freedom Δ		0	0	1	3

Note that optimal form values found in Column 4 all imply diminishing returns (to be called "damping" below), in accordance with Table 1. In particular, the optimal power value of the Frequency of service, close to the square root, means that additional service frequency yields progressively smaller gains. We will comment further below on the values of the BCT for Time and Fare, but note already that in this case the sensitivity to Time decreases slower than to Fare, as indicated by their respective BCT estimates (-0,05; -0,26), a situation where the marginal rate of substitution (the value of time) increases with Distance, as we will typically find in our survey.

The results of Table 2 (the published constants and their *t*-statistics are not reproduced) also show that, as soon as Box-Cox forms are used, regression coefficients loose all intuitive contents and require resort to elasticities to make some sense of the results. Consequently, the original paper (Gaudry & Wills, 1978) presented graphs and tables of the direct elasticities of modal shares with respect to network variables: we reproduce in Figure 2 the graphs of these elasticities (defined in Table 12), but not those of *t*-statistics, for the four modes.

The rail elasticities corresponding to model D of Table 2 are found just to the left of 0 on the abscissa of Figure 2.A for the Fare and Figure 2.B for the Travel Time. Changing the form from Linear to Optimal roughly doubles the rail Fare elasticity and halves the rail Time elasticity!

Clearly then, non linearity[20] would make a difference to HSR revenue forecasts as best fit direct modal elasticities dramatically and demonstrably differ from linear case values. Or do they really? Could discrete data re-establish the original battered linear case?

Figure 2. Direct Box-Cox Logit modal share elasticities (Canada, 92 city-pairs, 1972)

A. Absolute value of Fare elasticities

B. Absolute value of Time elasticities

---
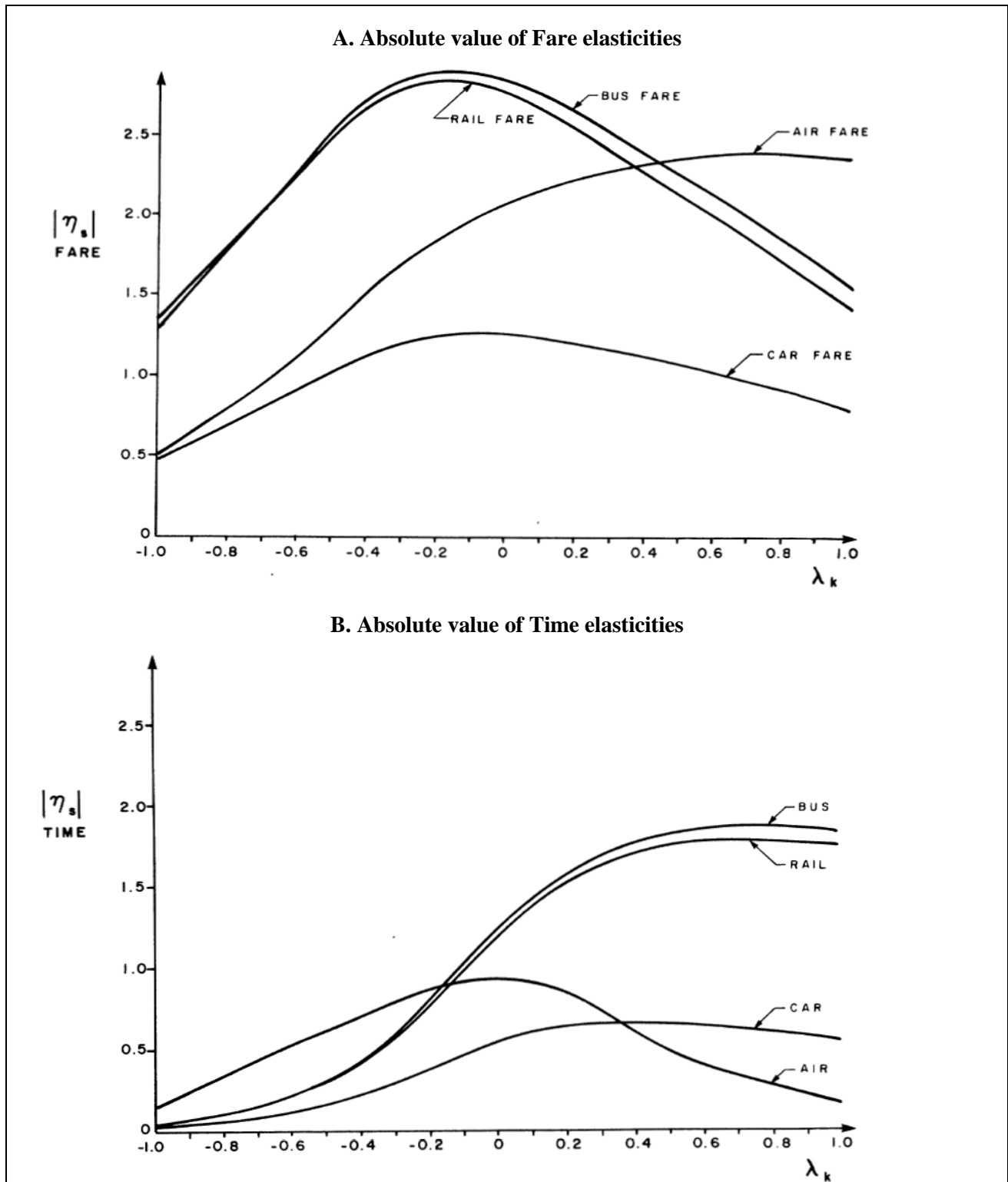
[20] The reader can consult similar graphs for corresponding *t*-statistics always computed for regressors conditionally upon the estimated values of the Box-Cox transformations: such conditional values are invariant to units of measurement of the transformed variables and do not suffer from the defects of unconditional estimates (Spitzer, 1984). The original paper graphs the evolution of conditional *t*-values of the Fare and Time variables against form values for the quantities shown in Figure 2. In this paper, all *t*-statistics presented are conditional on form.

3.3. A disaggregate model with multiple trip purposes, and more (1994)

As microeconomic purists balk at aggregate data and have long been hard to wean from the linearity of their Logit models (*e.g.* Morrison & Winston, 1985; Pickrell, 1987)[21], let us look in Table 3 at estimates of non linearity in a model based on a high quality Via Rail Canada database of 12 938 individual trips (4 402 business and 8 536 non-business) sampled in the Quebec City-Windsor Corridor in 1987 for the benefit of a multi-level government task force studying the HSR potential of this Corridor.

Is utility linear and separable? The original linear models by trip purpose found in Column 1 and in Column 4, originally specified independently by others (KPGM & Koppelman, 1990), are reproduced here from their re-analysis (Gaudry & Le Leyzour, 1994 or Tran & Gaudry, 2010a) in order to document in a continuous fashion first the impact of nonlinear forms (Column 2 and Column 5) and second the further impact of introducing in Modal utility functions (8-A) network variables pertaining to other modes, in the so called Generalized Box-Cox Logit form (Column 3 and Column 6) generalization of (6-B):

(8-A)
$$V_i = \beta_{i0} + \sum_n \beta_{in}^i X_n^{(\lambda_{in}^i)} + \sum_{j \neq i} \sum_n \beta_{in}^j X_n^{(\lambda_{in}^j)} + \sum_s \beta_{is} X_s^{(\lambda_{is})}$$
(Generalized Box-Cox)

(8-B)
$$V_i = \beta_{i0} + \sum_n \beta_{in}^i X_n^i + \sum_{j \neq i} \sum_n \beta_{in}^j X_n^j + \sum_s \beta_{is} X_s$$
(Generalized Linear)

where n and s are the indices for the network and socioeconomic variables respectively and the upper index refers to the mode, which can be seen as a first systematic attempt at giving a workable form to the Universal Logit idea (McFadden, 1975) because the characteristics of all modes are assumed to be relevant to the utility of each mode.

With suitable restrictions on the BCT values applied to own and cross network variables, this form (8-A) permits complex patterns of substitution and complementarity among modes, as classical demand systems naturally allow but the classical Linear Logit forbids by imposing a forced pattern of substitution based on an assumption of separability of utility among the modes. Clearly, parameters of (8-B), including all M constants, are all identified only if such utility functions are used in models like the LIN-IPT-L defined below in (12). There should in fact be two relevant questions: is utility linear? Is utility separable? These questions are also related: a linear form makes (8) collapse back to (5).

Non linearity, best fit and regression signs. Concerning the first question, the results shown in Table 3 indicate that: (i) the original linear model yields, for non-business trips, an incorrect sign on the travel cost variable (and consequently negative values of time) as evidenced in the darkly braided frames; (ii) as one introduces 4 Box-Cox transformations (one for each transport condition and one for Income), the gains in log likelihood are considerable[22] (from Column 1 to 2 and from Column 4 to 5) but incorrect signs of LOS variables are not fully corrected[23] until Column 6.

---

[21] These authors benefited from the high quality National Travel Survey carried out in the U.S. in 1977. The survey collected information on all trips of 100 miles or more made by members of 20 000 households covering the whole nation.

[22] One can show that the only transformation of the Frequency variable is not significantly different from 1.

[23] Clearly, as the results of Table 3 demonstrate in the spirit of Fridstrøm & Madslien (2002), sign change is a critical issue linked to how covariances among variables depend on their form. This means that models of apparently untested form are suspicious and that it is inadequate to justify linearity "for simplicity", as is still regularly done (*e.g.* Berry *et al.*, 2004). We come back to this issue in the last section of this paper. Results presented in Columns 4-6 differ very slightly from those found in Gaudry & Le Leyzour (1994) due to a recent correction of weights given to individual observations.

Table 3. Comparing Linear, Standard and Generalized Box-Cox elasticities and values of time

		Quebec City-Windsor Corridor of Canada 1987 (domestic intercity flows)					
Gaudry & Le Leyzour (1994)		**A. Business trips (4 402 observations)**					
Tran & Gaudry (2010a)					**B. Other trips (8 536 observations)**		
Weighted aggregate probability point elasticity[24]	*Column Case*	*1* *Linear*	*2* *Standard Box-Cox*	*3* *Generalized Box-Cox*	*4* *Linear*	*5* *Standard Box-Cox*	*6* *Generalized Box-Cox*
Original reference		*Model 3*	*Model 5*	*Model 6*	*Model 43*	*Model 48*	*Model 105*
Cost (access + in-vehicle)							
- Plane own cost		-0,53	-0,46	-0,12	*0,17*	-0,04	-0,08
- Train own cost		-0,05	-0,13	-0,07	*0,02*	-0,06	-0,06
- Bus own cost		-0,03	-0,13	-0,10	*0,05*	-0,17	-0,13
- Car own cost		-0,05	-0,10	-0,05	*0,01*	-0,07	-0,04
t-statistic of β_k		(-3,66)	(-7,23)	(-2,85)	*(2,14)*	(-6,91)	(-4,18)
Associated Box-Cox λ_k		**1,00**	**0,28**	**-0,22**	**1,00**	**0,48**	**-0,02**
Travel time (access + in-vehicle)							
- Plane own time		-0,32	-0,11	-0,21	-0,11	0,03	-0,00
cross w.r.t. car travel time		--	--	0,53	--	--	0,20
- Train own time		-0,18	-0,10	-0,12	-0,09	0,03	-0,02
cross w.r.t. car travel time		--	--	0,10	--	--	0,08
- Bus own time		-0,26	-0,14	-0,14	-0,23	0,07	-0,02
cross w.r.t. car travel time		--	--	0,13	--	--	0,12
- Car own time		-0,11	-0,05	-0,09	-0,06	0,02	-0,05
t-statistic of own β_k		(-10,19)	(-5,91)	(-4,31)	(-6,65)	(1,44)	(-1,51)
t-statistic of cross β_k		--	--	(4,59)	--	--	(8,38)
Associated own generic B-C λ_k		**1,00**	**1,80**	**0,61**	**1,00**	**0,47**	**4,81**
Associated cross generic B-C λ_k		**--**	**--**	**4,94**	**--**	**--**	**0,88**
Frequency							
- Plane own frequency		0,39	0,46	0,45	0,21	0,26	0,24
- Train own frequency		0,02	0,01	0,01	0,02	0,02	0,03
- Bus own frequency		0,12	0,10	0,11	0,13	0,17	0,17
t-statistic of β_k		(12,52)	(11,67)	(11,88)	(8,35)	(10,03)	(11,25)
Associated generic Box-Cox λ_k		**1,00**	**1,53**	**1,34**	**1,00**	**1,07**	**0,88**
Income (Gross Individual)							
- Plane (reference: car)		0,23	0,18	0,18	0,16	0,09	0,01
t-statistic of specific β_k		(4,99)	(5,00)	(5,05)	(4,01)	(2,72)	(-0,57)
- Train (reference: car)		-0,04	-0,04	-0,04	-0,02	-0,02	-0,00
t-statistic of specific β_k		(-2,61)	(-2,99)	(-3,00)	(-3,86)	(-5,09)	(-2,48)
- Bus (reference: car)		-0,10	-0,10	-0,10	-0,12	-0,15	-0,13
t-statistic of specific β_k		(-5,80)	(-6,89)	(-6,93)	(-13,13)	(-15,52)	(-12,31)
Associated generic Box-Cox λ_k		**1,00**	**0,16**	**0,18**	**1,00**	**0,33**	**0,44**
Other variables not reported		Trip origin in a large city [...]			Trip origin in a large city; party size [...]		
Log-likelihood		-1068	-1058	-1049	-2179	-2123	-2100
Degrees of freedom Δ		0	4	6	0	4	6
Hourly values of time (1987 \$)							
- Plane value of time		36,65	15,19	102,30	-23,39	-28,10	0,37
- Train value of time		36,65	7,44	16,68	-23,39	-2,23	1,12
- Bus value of time		36,65	5,33	7,42	-23,39	-1,81	0,47
- Car value of time		36,65	6,41	21,64	-23,39	-1,05	3,56

Clearly, it is inadequate to get rid of the problem by constructing a generalized cost variable for non-business trips, as was done by those who first analysed this high quality database with (nested or non nested) linear Logit forms (KPGM & Koppelman, 1990); and it hardly more commendable to entirely ignore results for this trip purpose (Bhat, 1995). The fact that the existence of statistical correlation depends on form, and conversely, goes deeper than mere fit: it is also the presence, sign

---

[24] The probability point elasticity is the usual elasticity of the probability multiplied by the probability of the alternative. For details, see Table 12 or Liem & Gaudry (1987, 1993).

and size of statistical correlation that non linear forms address and one should never be satisfied with correlations that are conditional on *a priori* form, irrespective of sign.

Accounting for non linearity and modal specificity: French "Price-Time" tandems. Even with non linear utility functions used flexibly and tested for mode-specificity, multinomial and nested Logit models based on separable Modal utility have limited ways of accounting for differing pair-specific substitution (different cross elasticities of modal choice) between HSR and other modes.

French authors (Abraham & Coquand, 1961), the first to link utility and multinomial choice, formulated as a Probit model but wisely approximated by a Logit model, applied it to a path choice problem some 10 years before work in the United States on a similar road path choice problem (Dial, 1971) and well before independent binomial road *tracé* choice work (McFadden 1968 or 1976a) or mode choice applications (Warner, 1962)[25]. In mode choice applications, French analysts have often used models in pairs (*e.g.* Air *vs* Rail and Road *vs* Rail) in order to better capture intermodal rail substitution specificity and adapt to the uneven availability of detailed path network data across the modes. In the absence of scheduled intercity bus services in France, only two models are necessary. This doubling up of Mode choice also opened the door to a doubling up of the Generation-distribution Total demand model, again to isolate train-specific induction effects that may differ from mode-abstract total induction effects.

But the Probit did not disappear in this context favouring use of models in pairs: a long favourite (*e.g.* Arduin, 1989, 1993), still in current application (Rail Concept, 2008), is a formulation where the value of time is assumed to be log-normally distributed within a binary Probit choice, an inflexible formulation that produces a particular asymmetry of the market response curve, in the spirit of Figure 1, without requiring the evaluation of multiple integrals.

But using multinomial Probit (MNP) instead of Logit (MNL) models, no more appealing to-day than in the time of Lisco (1967) six years after Abraham and Coquand had rejected it because of its multiple integrals, makes the issue of LOS form flexibility even less tractable, as pointed out by Bolduc (1999) [26]. However, if utility is not really separable because the Generalized Box-Cox Logit specification (8-A) applies, finer substitution (and potential complementarity) effects become feasible within the MNL and decomposition among pairs of modes looses much interest, especially with more than three modes. Nested Logit tree structures, that are not nested among themselves and are therefore hard to compare, also lose much appeal.

A workable form of the Universal Logit. Concerning the second question, note in Table 3 that the introduction of Car travel time in all utility functions increases considerably the log likelihood (from Column 2 to 3 and from Column 5 to 6), especially for non business trips[27]. This addition to the utility function certainly modifies the error correlation structure, providing a continuous alternative to non-nested hierarchies of Nested Logit tree structures, and reintroduces the possibility of complements, as demonstrated in the freight case (Gaudry *et al.*, 2008) to be used below[28].

---

[25] We do not underestimate the work by Theil (1969) on the link with information theory and the work on aggregate mode choice by Ellis & Rassam (1970) or Rassam *et al.* (1970, 1971) explicitly relating utility and Logit quantities.

[26] According to the author, "to implement a Box-Cox technique within a MNP setting represents a too formidable task". He goes on to dodge the flexible form issue by assuming lognormally distributed value of time, thereby implicitly obtaining market share response asymmetry with respect to Time.

[27] Fears of obtaining "incorrect" regression coefficient signs (Ben-Akiva, 1974) proved unfounded in Table 3, as a comparison of Columns 2 and 3 or 5 and 6 demonstrates, and in the freight example used for Figure 11.

[28] As noted, analysts interested in HSR demand may reintroduce specificity to cross-elasticities of mode choice with pairs of choice models (2 models if they have 3 modes). The Generalized Box-Cox Logit model not only reintroduces such texture but the possibility of complements, thereby reinserting the Logit in the family of demand systems.

It will also modify the variances of the error terms, that may or may not be homoskedastic[29] in linear space, a somewhat marginal problem –because only explanatory variables are transformed– to be addressed below as a general issue in the proper estimation of Box-Cox forms.

Also, the values of time differentiated by mode obtained in Table 3 (measured in 1987 Canadian dollars), make more sense for business trips in the optimal case of Column 2 than in the linear one of Column 1. It should also be noted that all Car travel time elasticities are strong and their *t*-statistics (in greyed cells) very high, which may be due to the addition of only a single off-diagonal term. If only one such term is chosen, time by car is surely the right one because the car, pervasive in that corridor (having almost 90% of the market in the total sample), is "the reference" obviously influencing the utility of all other modes. We also note in passing the clear presence of non linearity with respect to Income: indeed, why should marginal utility be constant with respect to income or to any other socio-economic factor if data really matter?

Again, form tests appear to make a difference that suggests not too much should be made of the difference between aggregate and disaggregate Logit models, an issue naturally raised early in linear models (Donnea, 1971, pp. 32-38; Kulash *et al.*, 1972): flexible BCT forms work on both. This should not come as a surprise: not only can the demand system generated by the Logit be associated with a single consumer maximizing a *deterministic* utility within a neoclassical framework (Anderson *et al.*, 1988) but the practice of fitting aggregate Logit models in the presence of consumer heterogeneity is theoretically justified "when all consumers are exposed to the same marketing mix variables and the brands are close substitutes" (Allenby & Rossi, 1991), *i.e.* in conditions that perfectly describe intermodal competition. Moreover, if such common factors are present at the micro level, nonlinearity is likely to remain in the aggregate[30] (Granger & Lee, 1999).

3.4. Another disaggregate model and HSR revenue maximization (1993-1994)

All above models were based on revealed preference (RP) data, whether aggregate or disaggregate. Would stated preference (SP) data originating from the Corridor, collected as usual without formally taking into account the possible non linearity[31] of the utility functions, change the answer?

Second thoughts again. Laferrière (1993b) and Ekbote & Laferrière (1994) used the same modelling sequence[32] as that just reported for the development of Table 3 models. They started from a data base built, and a nested linear Logit model specified, by others (SC Stormont, 1993) and tested the validity of the their linear forms by adding Box-Cox transformations on each of the cost and time variables, but not on the frequency of service variable, changing nothing else of the model originally specified independently from themselves. The results, found in Table 4, show the weighted aggregate elasticities of the HSR train mode choice probability with respect to Time and Cost under linear and optimal form specifications and the stunning gains in goodness-of-fit obtained by the application of BCT. The optimal BCT values all differ markedly from the linear case for both trip purposes: the gains are, if anything, clearer than with RP data[33]!

---

[29] Heteroskedasticity was studied by Bhat (1995) with this data base using only the business trip purpose data and dropping entirely the bus mode (3264 observations), a doubly arbitrary stance, in view of the fact that the bus has the most extensive network of all public services, that avoids facing the sign issue indicated in Columns 4-5 of Table 3 for non business trips and perhaps also unreported problems in the treatment of heteroskedasticity of bus utility functions.

[30] The issue of temporal aggregation, also treated by Monte Carlo simulation in that paper, is distinct from that of aggregation over agents. The later literature generally shows that weak conditions are required to obtain well behaved aggregate functions, aggregated on income (Hildenbrand, 1983) or preferences (Grandmont, 1992), such as the negative definiteness of the matrix of uncompensated price effects, for which tests exist (Korenman *et al.*, 1988).

[31] In designing samples, the SP data set should be created with a view to obtaining observations that are orthogonal not in linear space but in the optimal non linear utility space. How to best do this is a research issue.

[32] Described in full in Laferrière (1993a).

[33] Not surprisingly, some authors (*e.g.* Lerman & Louvière, 1978) have suggested a two-step approach whereby results from experimental data on functional forms derived by modifying each variable in turn (a technique called functional

After correcting the inherited linear form, the authors' next step consisted in finding the most profitable money and time combination in the Corridor. Formal revenue maximization over Cost and Speed (in a range from 120 to 400 km per hour) was performed over various Frequency assumptions and rail ridership summed over all OD pairs (as well as specific OD pairs) using an updated version of the database (used in Table 3) produced by the government task force (OQRTTF, 1991) and the pivot method (Laferrière, 1994).

Although their actual procedure is somewhat more complex than just described, in that it also involved adaptation of the modal constants to reflect revised sampling rates, some very interesting lessons can be learned from their use of non linear forms.

Table 4. Model results (SP data, 1991) and revenue maximization (RP data, 1992) for 2010

A. Regression	Business trips		Non business trips	
Column	*1*	*2*	*3*	*4*
(Data: 1991) *Case*	*Linear*	*Non linear*	*Linear*	*Non linear*
Train mode split elasticities*				
Cost	-0.139	-0.315	-0.169	-0.192
Travel time	-0.595	-0.502	-0.067	-0.134
Box-Cox transformations				
Cost	1.000	**0.256**	1.000	**0.312**
Travel time	1.000	**0.562**	1.000	**-0.102**
Access time	1.000	**0.194**	1.000	**0.513**
Goodness-of-fit				
Log likelihood	-5182.03	-4898.75	-8730.73	-7909.65
Degrees of freed. Δ	0	3	0	3
B. Percent revenue gain per 1% speed increase in market	**Business trips** (Non linear case)		**Non business trips** (Non linear case)	
	Quebec City Montreal	Montreal Toronto	Toronto Windsor	Quebec City Montreal	Montreal Toronto	Toronto Windsor
120 to 200 km/h	1.62	1.41	1.19	1.07	0.65	0.85
200 to 300 km/h	1.19	1.17	0.88	1.06	0.67	0.89
300 to 400 km/h	1.00	0.95	0.71	1.07	0.66	0.89
Road Distance (km)	270	560	330	270	560	330
C. Total forecast	**300 km/h scenario revenue maximizing linear and non-linear results**			
(Data: 1992) *Case*	*Linear*		*Non linear*	
Trip Forecast	5,83 million passengers		7,70 million passengers	
Revenue Forecast	$ 576 million		$ 446 million	

* Elasticities for other modes are found in Laferrière (1993b, Transparency 5). For their definition, see Table 12.
** Calculated for 2010 network with RP sample data of 1987 updated in 1989 and 1992.

First, it is clear that, as shown in Part B of Table 4, strongly increasing marginal revenue gains at relatively low speeds progressively decrease with higher speeds (affecting In-vehicle time proportionately) in the three main business trip markets. Also, in the non-business trip market, there is a big difference between constant returns to Speed in the Quebec City-Montreal market and decreasing returns in the other two markets. Second, Part C indicates that the dominant best fit non linear forms implied lower HSR revenues than those obtained with the (non optimal) linear form of the same models despite higher revenue maximizing passenger levels. Air Canada and CP Rail have apparently never considered HSR investments again.

---

measurement) are used as maintained, or even final, hypotheses for the same specification estimated with RP data. This is a good idea when collinearity of the RP data is high, but the functional measurements have form problems of their own in that the *ceteris paribus* procedure applied to each variable in turn does not imply their independence (the orthogonality of their distributions). Ekbote and Laferrière follow an analogous sequence: they obtain first stage form results from BCT estimates on SP data and apply them (without re-estimation) to the RP Via Rail Canada 1987 database (used in Table 3), updated to 1992.

Shifting the burden of proof? Perhaps the burden of proof, on both the marginal constancy and the separability of representative Logit utility functions, should be reversed and the linear case never assumed unless it can be demonstrated to be correct.

As the Box-Cox transformation effects a local approximation of form, it is quite possible that, in some problems where observations pertain to relatively small domains (or for other reasons discussed in Section 5), the linear form be sufficient. But, until tests of more global approximations, such as those carried out with Fourier transforms (*e.g.* Gallant, 1981) reach Logit models, the use of Box-Cox specifications seems warranted for HSR cases where non marginal variations in Travel Time are always required.

What other objections have been raised to the use of endogenous forms? We now consider two, addressed with Swedish models. As analysts often decompose markets among segments, and not just among trip purposes, the first objection is that the use of non linear forms might confuse form with market segment differences. The natural extension of that objection, related to the surge of "Mixed Logit" models generalizing taste differences among individual, is also addressed. The second frequent objection is that form modification is an indirect way of stabilizing the variance of errors that do not have a constant error variance under linearity assumptions, *i.e.* that suffer from heteroskedasticity under a linear formulation. This second objection pertains more generally to the effect on form estimates of nonspherical distributions of the random terms that constitute the stochastic, as opposed to the "fixed" part, of models.

Swedish studies on the relevance of these two objections are of particular interest because Sweden, like Canada, allows scheduled intercity bus services and does not regulate them out of existence to protect the railways, as do France and Germany[34]. Sweden is therefore a 4-mode country where buses provide an important intercity service, if not the first public supply in terms of network extension (points served), as in North America. The Swedish attitude to model development also makes it relevant to our concerns.

---

[34] Severely limiting scheduled national and international bus lines but allowing chartered buses, significant for tourism: it is estimated that there are about 2000 chartered German buses in Paris on a typical week-end.

4. Challenges to non linearity: market segments and stochastic terms

4.1. Form and segmentation lessons from Sweden

In 1979, the Swedish Board of Transport (TPR) initiated a policy of promotion of intercity modelling in order to obtain analyses and forecasts of all domestic passenger transport services. At first, many of the models were aggregate, but discrete choice models were progressively developed.

Count data generation combined with aggregate Nested Logit mode choice. Among the aggregate models, one finds for instance a potentially interesting 3-mode model (the bus unfortunately neglected) of interregional flows among 70 regions of Sweden formulated by Sävenstedt & Uhlin (1985). They multiplied a Count data component for trip generation, estimated by Poisson regression, by an aggregate[35] nested (or "structured") Logit mode and destination choice model component, estimated by a non-iterated version[36] of the Berkson-Theil estimator weighted to take a form of heteroskedasticity into account. This is *prima facie* an exceptional combination to learn from, but it may be even more instructive to describe the efforts then made to introduce some non linearity in the behavioral response to the Frequency of service.

All network variables appeared linearly in the mode choice model but modal Frequency G_m was specified as $ln[1-exp(\beta_d\ FR_m)]$, where by assumption $\beta_d < 0$ and FR_m, the day-time departure frequency[37] for the public modes, was assumed equal to infinity for the car. There are two aspects to this complexity: the issue of form and the problem of car frequency. To capture the former, consider the first and second derivatives of the representative utility functions with respect to Frequency so defined, namely:

$$(9\text{-A}) \qquad \frac{\partial U_m}{\partial FR_m} = \beta_{mk} \left\{ -\frac{\beta_d[exp(\beta_d FR_m)]}{\{1-[exp(\beta_d FR_m)]\}} \right\}, \qquad \text{and} \qquad \frac{\partial U_m^2}{\partial FR_m^2} = \beta_{mk} \left\{ -\frac{\beta_d^2[exp(\beta_d FR_m)]}{\{1-[exp(\beta_d FR_m)]\}^2} \right\},$$

$$\text{where} \qquad\qquad \{ > 0 \} \qquad\qquad \text{and} \qquad\qquad \{ < 0 \}.$$

Because the term $[exp(\beta_d\ FR_m)]$ in (9-A) is always a positive fraction, the bracketed expressions multiplying the regression coefficients β_{mk} are positive and negative, respectively: consequently, increased departure frequency always yields positive marginal utility but by progressively smaller amounts, exactly as happens with successive increases of X_{mk} in Table 1 if and when $\lambda_{mk} < 1$.

The specification of car frequency poses a distinct sub-problem because it is natural to give it a high and incorrect value (as inappropriately done sometimes) or to ignore it altogether as a variable (as correctly done here). The desirable path in the estimation of Frequency coefficient β_{mk}, adopted by the authors and in the models found in Table 5 or in the set of Tables 7-10 and 18, consists in defining Modal utility functions in a general way for all modes, for instance in a linear example (and neglecting the Fare, *etc.*), as:

---

[35] Aggregate Nested Logit models are extremely rare in transportation, much more than aggregate Logit models.

[36] Unfortunately, this procedure, based on applying Ordinary Least Squares (OLS) to the logarithm of the odds of making a particular choice relative to the choice of a reference alternative, yields β_k coefficients that depend on the reference alternative chosen if the (weighted) OLS estimator is not iterated until convergence to the maximum likelihood value using the Σ variance-covariance matrix, of dimension (M-1)x(M-1), among the residuals of the resulting equations, as shown by Wills (1982): this estimator has the known form $\beta = (X'\Sigma^{-1}X)^{-1}X'\ \Sigma^{-1}y$. Their one-pass approach, whereby analysts incorrectly think that the coefficients of their specified mode choice model are estimated independently from the reference mode chosen, is still frequently encountered, for instance in a model of the choice of port on the Atlantic seaboard, called "CPB-VITO" (Veldman & Bückmann, 2003; ECORYS Transport *et alii*, 2004).

[37] As it is assumed that $FR_{car} = \infty$, so that $G_{car} = 0$, we conclude that the actual relevant regression term is of the form $[\gamma_{frequency}\ ln(G_m)$, for $m \neq car]$. In this case, the null car frequency disappears as a regressor and implicitly modifies the car representative utility constant, as shown in the linear example (9-B).

(9-B) $\qquad V_m = \left[\beta_{0,m}\, C_0\right] + \beta_{\text{Time},m}\, \text{Time}_m + \beta_{\text{Frequency1},m}\, (1-Z)\, \text{Frequency}_m + \beta_{\text{Frequency2},m}\, (Z) \!\!\!\!\lrcorner$

with $Z = 1$ for Car and 0 for public modes (and $m = 1, \ldots, M$), the unbiased estimation of which simply requires a car specific constant, as the following explicitation makes clear:

(9-C) $\qquad V_{\text{Car}} = \left[\beta_{0,\text{Car}}\, C_0 + \beta_{\text{Frequency2},m}\, 1\right] + \beta_{\text{Time},m}\, \text{Time}_m ,$

(9-D) $\qquad V_{\text{Other}} = \qquad \left[\beta_{0,\text{Other}}\, C_0\right] \qquad + \beta_{\text{Time},m}\, \text{Time}_m + \beta_{\text{Frequency1},m}\, \text{Frequency}_m .$

Two questions then come to mind: (i) would not parsimony argue for the adoption of a Box-Cox transformation of Frequency, perhaps with a fixed *a priori* power parameter value? (ii) if the marginal utility of Frequency was assumed to vary with the level of that variable, why was it assumed to be constant for the other network variables? Such practical questions were "in the air" at the start of work on discrete choice models (Algers, 1984) made possible by a remarkable series of Swedish surveys.

National Travel Surveys, a Swedish ritual. Concerning disaggregate models for long-distance travel, the very important stream of national Swedish discrete choice models started with the availability of the 1984/85 National Travel Survey, containing about 1800 private and 500 business tours, which made possible a first intercity model (Algers, 1993). The subsequent 1994-97 survey, bringing forth a much larger data base — about 10 000 observations on private trips and 3 500 business trips —, allowed for a second demand model built as part of the Swedish national travel demand model system SAMPERS (Beser & Algers, 2001) spanning 670 domestic and 200 foreign zones and in use by the Swedish transport planning authorities after 1998. A still larger data set based on the 1994-2000 survey (notably including about 15 000 observations on private trips) was used in 2003-2004 for an update activity of the SAMPERS system, including its demand model component (Algers, 2004a). A fourth model, based on the 2005-06 survey, and specifically designed for HSR potential demand determination, is in progress[38] and, in view of the first results (Algers, 2011), should be completed in 2011.

The first three "official" milestones of Swedish intercity disaggregate demand model development all consisted in linear nested logit models for mode, destination and trip frequency choice. There is one exception to this, namely the form of the Headway parameter. As was found in previous modelling (Algers & Gaudry 1994) and in the 1994 Swedish Stated Choice Value of Time study, Headway has a decreasing marginal disutility. This was taken into account in the second and third milestone models by a piecewise linear headway formulation for that variable, in which the relative values of the pieces were constrained to those found in the 1994 Value of Time study (these relations are also found to hold in the 2008 Swedish Value of Time study). In addition to the non linear tests made on the first milestone model, a number of non linear form test variants were also developed for the third milestone models.

The first effort, carried out at the Royal Institute of Technology (Algers & Gaudry, 1994), was based on the model estimated with 1984/85 survey data and focussed on the links among form, segmentation and heteroskedasticity. The second effort some 10 years later, as reported by the principal researcher (Algers, 2004b), was carried out on commission of the planning authorities in connection with the SAMPERS model update: it featured both modal nest and form sensitivity tests on the combined mode and destination model, but only for the main private travel purpose.

We now consider these two waves of form tests in reverse chronological order, addressing statistical heteroskedasticity issues last.

---

[38] Author's communication with the Staffan Algers of the Royal Institute of Technology who kindly authorized the use of Figure 2 and vetted this Section of the paper.

A. Form tests in a model of private trips (2003)

As noted above in the Quebec-Windsor Corridor case documented in Table 3, it is often more difficult to model private trips than business trips, and all the more critical to do so that non-work trips constitute an increasing proportion of total trips in advanced continental European air markets where they often reached 50% of total air trips in the 1990's. Still, form sensitivity tests of a model for such trips yielded the unambiguous results graphed in Figure 3.

Those sensitivity tests, carried out in 2003 (Transek, 2003; Algers, 2004b) on a combined mode and destination choice model for the main private travel purpose with the 15 000 observations available for the SAMPERS update, only used generic Box-Cox transformations on the Cost and In-vehicle time variables, the base model (itself a variant of the updated milestone with respect to mode nests) already including a piecewise linear transformation on Wait time (defined as half Headway). As a result, the log-likelihood value increased from -26950 in the linear specification to -26512 (with a difference of 2 degrees of freedom), a massive gain of 438 units, seen on Figure 3, taken from Algers (2004b), where two local maxima occur[39]. As these local maxima were found in the tested range between 0 and 1 with coarse grid[40] steps of 0.20, the global maximum had to be pinpointed by a finer grid search with steps of 0.05 indicating an optimal lambda combination of 0.40 for cost and 0.15 for in-vehicle time.

Figure 3. BCT tests on Cost and In-vehicle time, SAMPERS mode & destination choice model

---
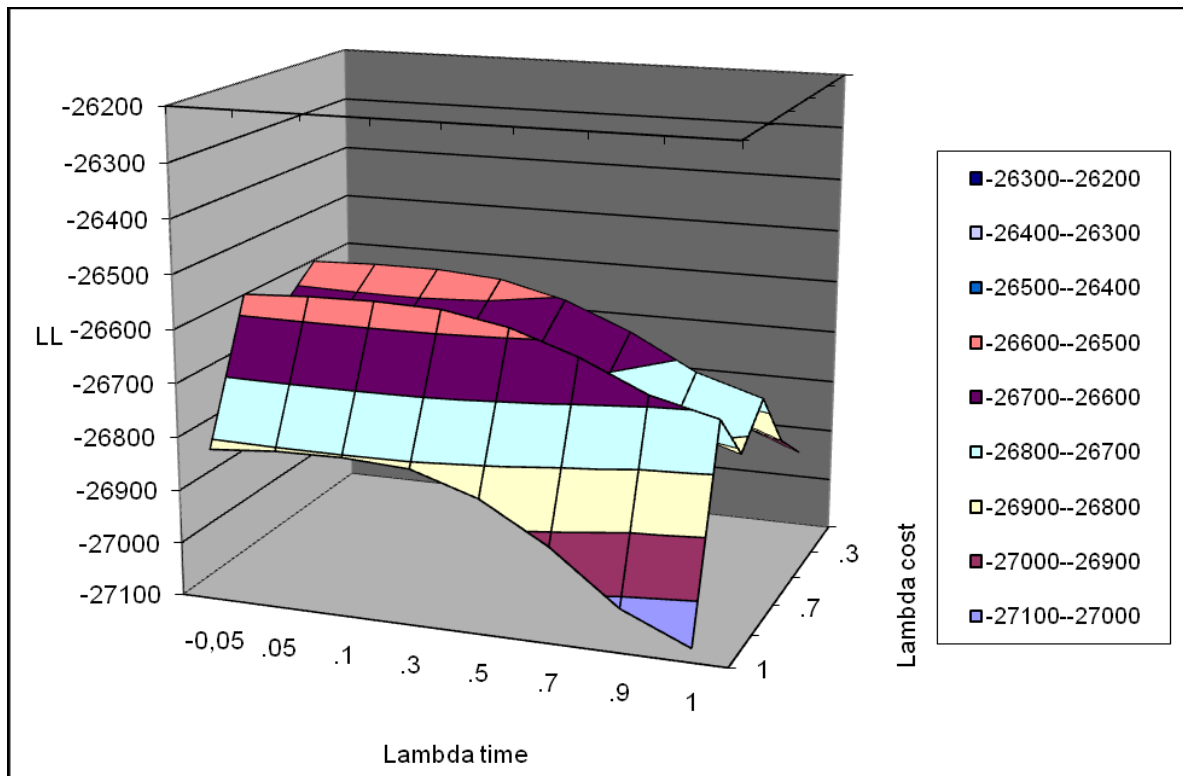
[39] The current version was generously contributed by Algers.

[40] In a mode and destination choice model where the mode choice components (the denominators) give rise to *log sum* terms used to explain destination choice, a grid can be performed by transforming the Cost and Time input variables and finding the corresponding log likelihood values. As the *t*-statistics of the transformed variable regression coefficients will *per force* be obtained conditionally upon the hypothesized values of the transformations, they are invariant to changes in the scale of those variables and automatically meet Spitzer's (1984) objections to unconditional estimates.

B. Form and segmentation in models for business and private trips (1994)

We noted already that the possibility of a substitution between form and market segmentation might raise doubts about the validity of such impressive results on the non linearity of utility and on results from eventual tests of non separability of Logit modal utilities as well.

Market segmentation, the art of classifying a population into heterogeneous groups of homogeneous members, *i.e.* of accounting for the existence of subpopulations in samples, remained disjoint from functional form analysis for a surprisingly long time despite the fact that the role of segmentation has long been recognized in transportation demand analysis in the context of fixed (linear) forms (Hensher, 1976). Their linkage, despite its obvious appeal, was apparently explored systematically for the first time only in Algers & Gaudry (1994) as a variant of the first Swedish milestone model.

Considering base model business and non business trip purpose specifications, these authors reported on the effect of segmentation criteria related as closely as possible to Income, often assumed to hide strong heterogeneity of preferences. For business trips, the assumption was, in the absence of data on the Income of individuals, that underlying preferences might be related to their Full Time Salaried Employee (*vs* Part-time or Independent worker) status; for private trips, the availability of Household Income in 95% of the sample argued, despite the reduction in the number of available observations, for a distinction between High (*vs* Low) household Income membership. The most conservative of their results, reported in Table 5 as answers to a question, require comment.

When a variable, say Wait time, is segmented into components H and L as a function of the sampled individual's household Income, two regression coefficients β_H and β_L replace the previous β_{HL} singleton but there naturally also exists two ways to specify the non linearity parameters of the two complementary Wait time segment variables: they can be transformed by a common Box-Cox power λ_{HL} or by segment-specific powers λ_H and λ_L. In the latter case, other conditions additionally need to be met because the segments now contain subsets of zero-value observations (or at least of "holes") which require a corresponding dummy variable to guarantee the invariance of the power parameter estimates to changes in units of measurement of the segmented component variables.

Table 5. Non linearity *vs* income-related segmentation in the first milestone model for Sweden

Algers & Gaudry (1994)		Should *non linearity AND segmentation* be specified for		
Variable segmented or transformed		the Business trip model?		the Private trip model?
1	**Cost**	*No*	include either, but not both: they are substitutes	*Yes* · include both: they are independent
2	**In-vehicle + Transfer time**	*No*	include either, but not both: they are substitutes	
3	**In-vehicle time**			*Iff* · cost is *non linear*, then segment on time
4	**Wait time**	*No*	include *only non linearity*	*No* · include neither
Number of observations from 1984/85 survey		*534*		*1758*

The choice is in practice between two formulations involving only one dummy variable because observations belonging to segments constitute complementary and disjoint subsets. It is between:

(10-A) $(\beta_H , \beta_L , \lambda_{HL})$ [common segment curvature specification], and

(10-B) $(\beta_H , \beta_L , \lambda_H , \lambda_L , \beta_{DUMMY})$ [specific segment curvature specification].

And it should be clear that (10-A) is more conservative than (10-B) because segmentation in principle applies as much to the curvature of the segment as it does to the associated segment regression coefficient. Although both specifications were tried for most of the variables of interest

in Table 5 (considered one at the time and jointly), the exhibited results, based on the best performing (joint) models in terms of log likelihood drawn only from (10-A) trials, are, in that sense, very conservative because form powers could naturally also be segment specific.

Broadly, they indicate that segment-independent non linearity, detected by a Box-Cox transformation of a particular variable, and segmentation of that variable can be substitutes, as one would spontaneously expect from segmentations on say Distance related variables, but can also be complements, or even independent dimensions of model specification. At least for Cost and Wait time variables, there exists in these Swedish data "inherent" non linearity to be taken into account and the remaining issue is really whether non linearity might even be segment-specific.

But does this argument carry over to the fragmentation of segmentation to the extreme extent of assuming that taste heterogeneity can be represented by a full distribution of individual weights for a given variable?

C. Form and the randomisation of regression coefficients

A challenge to the existence of non linearity is sometimes thought to be posed by the treatment of coefficients of the utility function as random instead of fixed, an old innovation in classical models (Swamy, 1970; Johnson, 1977, 1978), including transportation (Hensher & Johnson, 1979a) where the link with form was made early by Johnson (1979)[41], recently taken up in Logit kernels under the « Mixed » Logit label, a formulation thought capable of approximating any random utility model (*e.g.* McFadden & Train, 2000). But do Mixed Logit models need form estimates?

The atomisation of segments effected by randomisation of regression coefficients poses problems because the distributions of coefficients are unknown. Should the regression coefficients of Income, itself often log-normally distributed, follow an unbounded, or a doubly censored, normal distribution? Does the distribution of a gender variable follow a particular law with a regression component related to testosterone or estradiol, or to both ─ appearing as levels, ratios, or both with BCT? In Mixed Logit models, the information matrix[42] does not have a closed form, which implies an undefined efficiency bound (Cirillo, 2005). Despite these design handicaps, Lapparent *et al.* (2009) have shown with long-distance data on three countries that BCT on Time, Cost and Access time were typically different from both zero and one in a Mixed Logit model.

More importantly, the form objection can be turned on its head: Orro *et al.* (2005) have demonstrated with Box-Cox Mixed Logit model simulations (using two BCT, on Fare and Travel time) that the recent popularity of the Multinomial Mixed Logit may well be due to the fact that the true relationships are not linear and should have their curvature estimated rather than postulated, as many micro-economists might have long suspected: perhaps many specifications are more mixed up that mixed. Indeed, to extend (10-B), if β_k regression coefficients vary across individuals due to taste heterogeneity, why would their marginal utility trade-offs all be linear? Logically, should not the marginal utility across individuals be randomised with Box-Cox form parameters obtaining distributions as well, rather than be assumed constant, as demonstrated again by Orro *et al.* (2010)?

In fact, concerning this attitude, one might guess that some other model parameters also determined jointly by estimation might have even better claims to distribution parameterisation than taste or form parameters: serial autocorrelation parameters, for instance, are likely to differ enormously

---

[41] The author was carefully pessimistic, at least in a technical sense, about prospects for this combination: "Unless more detailed results for the Murthy estimator are encouraging, it may be too much to expect the Box-Cox transform to work in more complex random coefficient models" (p. 1035). Murthy (1976) had suggested an estimator for the Hildreth-Houck (1968) linear model with random coefficients but had provided no empirical estimates.

[42] This important point concerning the matrix of expected values of the second derivatives of the Log Likelihood function was brought to our attention by Lasse Fridstrøm.

among individuals as they can capture stable idiosyncrasies that are notoriously difficult to account for with regressors, such as different discount rates. But inherent model misspecification brings us naturally to the modelling of the stochastic part of the specification.

4.2. Form and non spherical distributions of residuals

If segmentation and form are indeed distinct modelling dimensions, what else might present a confounding influence on form? Could form estimation then compensate for distributions of residuals violating statistical requirements of independence and constancy of variance?

In both model classes defined by (6-A) and (6-B), the stated assumption of independence matters in principle as much as that of error variance homogeneity, but our emphasis will be on the latter even if both have a critical role to play in the determination of the standard errors of all parameter estimates, and hence on the reliability of their *t*-statistics. Unfortunately, the usual econometric proofs of bias or inconsistency of estimates obtained when these conditions are not met typically assume that the "fixed" part of the postulated model is correct.

A. The diagnostic and correction of error interdependence and variance heterogeneity

In reality, models are incorrect and both non homogeneity of error variance and interdependence among errors are more caused by model misspecification, and notably by the absence of relevant regressors, than by pure randomness. If for instance the true model is "dynamic" but has been specified as a static model, correcting for serial autocorrelation can indirectly improve the original formulation of the model, a result (Spanos, 1987-88) that should also hold, *mutatis mutandis*, for corrections of spatially autocorrelated errors and for error variance function corrections, such as $[f(Z)_t]^{1/2}$ for (6-A) and $\mu_i[f_i(Z_i)_n]^{1/2}$ for (6-B), further discussed below.

In practice, however, meaningful models of variance determination are notoriously difficult to formulate intuitively, as the representation of our ignorance is inherently limited, in contrast to models of interdependence that are readily driven by the mechanics of time or space indices. In consequence, to this day models of heteroskedasticity remain very much the minority in the literature relative to models of interdependence. This low share hides some systematic differences between linear multivariate[43] regression model classes, as implied by candidates to "first of" status in Table 6 where spatial autocorrelation (to be addressed superficially in Appendix D) is neglected.

(i) formal testing for the presence of error interdependence or of error variance heterogeneity always came after *ad hoc* corrections in all Classical and Logit cases, if it came at all;

(ii) *ad hoc* corrections for error interdependence came earlier (1949) than for error variance heterogeneity in Classical models (1961) but the reverse holds in Logit models where error variance[44] corrections came late (1970) and began even later for interdependence, apparently at a moment of the 1980's that is difficult to identify precisely.

Table 6. Correcting nonspherical distributions of residuals in linear multivariate regression

	Serial interdependence		Nonhomogeneity of variance	
	First *ad hoc* corrections	**First tests of presence**	**First *ad hoc* corrections**	**First tests of presence**
CLASSICAL	Cochrane & Orcutt (1949)	Durbin & Watson (1950) Durbin & Watson (1951)	Anscombe (1961)	Levene (1960) Goldfeld & Quandt (1965)
LOGIT	Gaudry & Wills (1979)[45].	mid-1980's	Cox (1970)	Davidson & MacKinnon (1984)

---

[43] The seminal Generalized Least Squares idea attributed to Aitken (1934-1935) did not deal with a multivariate case.

[44] The presence of heteroskedasticity in binary cases is easy to intuit because, with observations coded 0 or 1, the error term is necessarily [0 - $\mathbf{p_m}$*] or [1 - $\mathbf{p_m}$*], where $\mathbf{p_m}$* denotes the probability calculated by the model. Cox's (1970) recipe of adding ½ to the choice probabilities, taken up by Domencich & McFadden (1975, p. 109), failed to spread.

[45] Aggregate Logit.

Note in passing that some forms of heteroskedasticity correction can deeply modify the own and cross-elasticities. In (6-A) the presence of heteroskedasticity will, independently from the presence of autocorrelation, add a complex second term to the usual elasticity of y with respect to X_k (evaluated at samples means) $\tilde{\eta}(T_{TOT}, X_k) = T_{TOT}^{-\lambda_{TOT}}(\beta_k X_k^{\lambda_{X_k}})$ if the X_k regressor of interest also belongs to the heteroskedasticity correction function[46]:

(11-A)

$$\tilde{\eta}(T_{TOT,t}, X_{k,t}) = \frac{1}{T_{TOT,t}^{\lambda_{TOT}}} \beta_k X_{k,t}^{\lambda_{X_k}} + \left\{\frac{1}{2}\delta_m Z_{m,t}^{\lambda_{Z_m}}\left[T_{TOT,t}^{(\lambda_T)} - \sum_k \beta_k X_{k,t}^{(\lambda_{X_k})}\right]\right\},$$

$$\text{if } f(Z)_t = \exp\left[\sum_m \delta_m Z_{m,t}^{(\lambda_{Z_m})}\right] \text{ and } f(Z_t) \text{ contains one } Z_{m,t} = X_{k,t} .$$

In (6-B), for instance, mode-specific scale factors in $\mu_i[f_i(Z_i)_n]^{1/2}$ will modify the consistency with IIA of the original structure (4)-(5) because, even if $[f_i(Z_i)_n]^{1/2} = 1$, the cross elasticity between a probability $p(i)$ and a variable belonging to any utility function V_j will now depend on the μ_j, as in:

(11-B)

$$\eta(P_i, X_{jk}) = \beta_{X_{jk}} X_{jk}^{\lambda_{jk}}\left(-P_j\right), \quad \text{if } \mu_j = 1,$$

$$\eta(P_i, X_{jk}) = \beta_{X_{jk}} \frac{X_{jk}^{\lambda_{jk}}}{\mu_j}\left(-P_j\right), \quad \text{if } \mu_j \neq 1 .$$

In single-equation case (6-A)-(6-D), heteroskedasticity is defined as $\{[f(Z)_t] \neq 1\}$. But the definition is more complicated in multiple-function case (6-B)-(6-E) where it may occur either across the observations pertaining to any given utility function equation, *i.e.* $\{[f(Z_i)_n] \neq 1$ for any $i\}$, or across scale factors of the M equations $\{i.\ e.\ [\mu_i \neq \mu_j]$ for any i and $j\}$, or even across both observations and equations together. Note that setting all $[f_i(Z_i)_n]$ and all μ_i equal to 1 corresponds to the so-called "standard Weibull" (Johnson & Kotz, 1970, p. 253) implicitly homoskedastic case.

As the correction effected to obtain a constant variance requires dividing the regressand and all regressors, including the constant, by $[f(Z)_t]^{1/2}$ in the Classical case and by $\mu_i[f_i(Z_i)_n]^{1/2}$ for each representative utility function in the Logit case, correcting for heteroskedasticity always involves increasing the relative importance of (homoskedastic error) randomness in the specification, a fundamental matter in both Classical and Logit structures. It also involves specific risks of creating numerical outliers and collinearity to the extent that the division even by a strictly positive $[f(Z)]^{1/2}$ that avoids negative variances can sometimes create a relatively dominant observation — if the chosen specification for instance includes a Z_H variable raised to a positive power.

B. The presumed link between form of variables and variance stabilization

To see why such a minority issue as heteroskedasticity could specifically matter here, note that maximizing Likelihood functions notably involves minimizing a function of calculated errors, such as sums of squares for (6-A)[47].

Classical regression. In this case, modifying the BCT of the dependent variable y in (6-A) directly affects the size of errors. Transforming the dependent variable is therefore quite different from transforming independent variables (Davidson & MacKinnon, 1993). For this reason, analysts have long used logarithmic transformations to obtain residuals of smaller variability than those obtained if y were linear, implicitly assuming that Case 1 of Figure 4 held for their problem rather than Case 2 or 3.

---

[46] The detailed proof may be found in Tran & Gaudry (2010a).
[47] If we ignore the Jacobian of the transformation of the dependent variable that appears in the likelihood function of y.

Figure 4. Influence of a logarithmic transformation on error variance

Case 1: Reducing both nonlinearity and heteroskedasticity

Case 2: Reducing nonlinearity and inducing heteroskedasticity

Case 3: Reducing heteroskedasticity but inducing nonlinearity
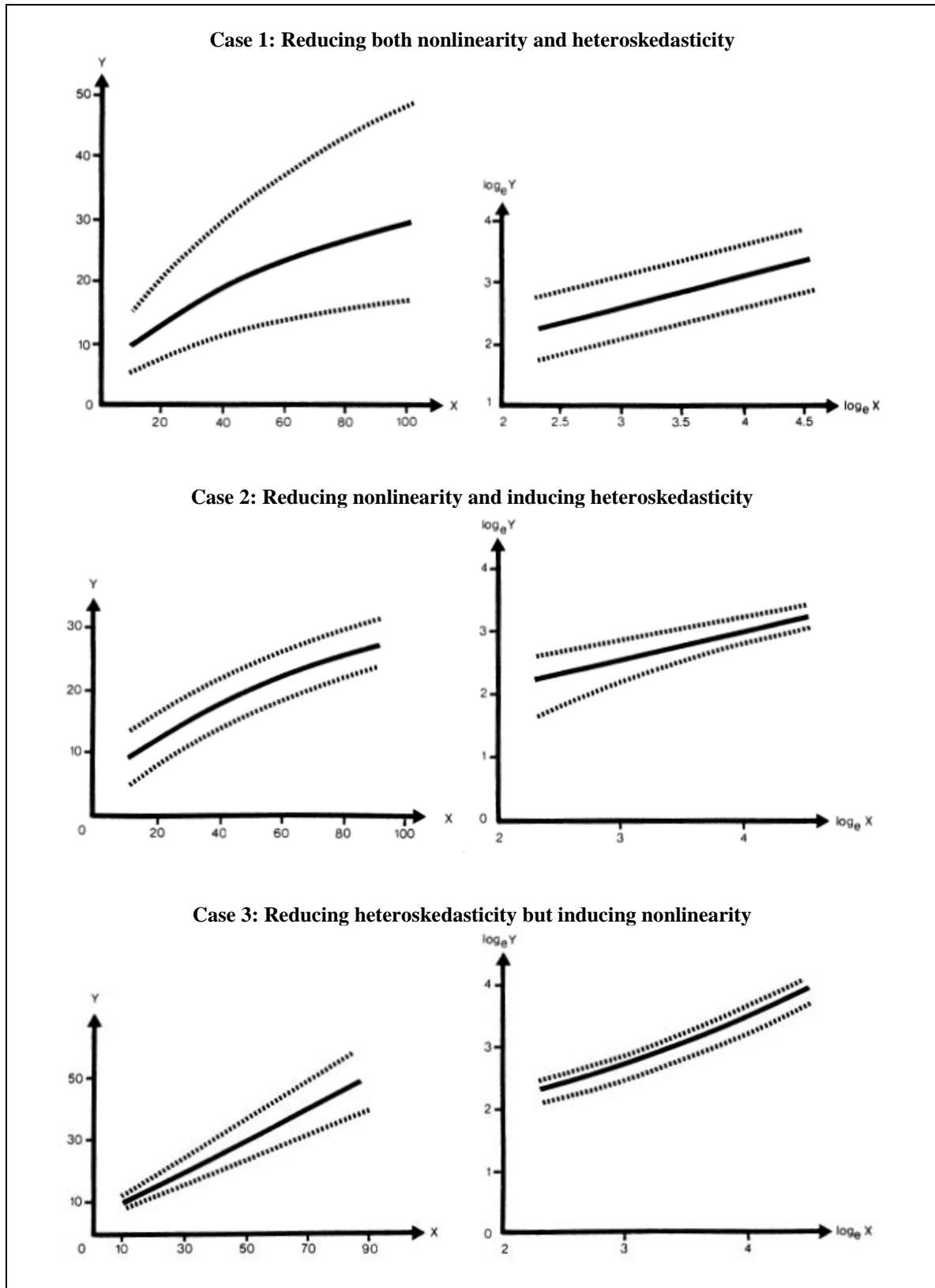
That figure illustrates that the logarithmic transformation, or more generally the link between nonlinearity of *y* and heteroskedasticity, is not simple because power transformations can reduce, as well as increase, error variance, and this without even considering BCT powers larger than 1. On

the same lines, it is sometimes erroneously believed that the BCT applied to *y* is always a variance[48] stabilizing transformation, and never the opposite. It would be more helpful to recognize that a BCT transformation on *y* necessarily affects both nonlinearity and error variance, as demonstrated in Figure 4, and that a second handle is required if the twin modelling targets of due model form estimation and simultaneous error variance stabilization are to be met. For that purpose, the structure of heteroskedasticity should be estimated jointly with the form of the model, for instance as specified[49] in (6-D) and implemented in many documented and publicly available algorithms (Tran & Gaudry, 2008c; Tran *et al.*, 2008). Two targets require two control instruments.

Having one instrument per target[50] may be more satisfactory than estimating BCT forms first and testing afterwards for their sensitivity to heteroskedasticity, as proposed earlier by Zarembka (1974), but some wisdom in application is still warranted because all regression variables are not equally linked to error variance. Although the use of BCT on any explanatory variables of a model can in principle have similar variance stabilization or destabilization effects as transformations of *y*, such effects will be mooted by the compensating regression coefficients[51].

Because of this tempering property of coefficients not constrained to 1 like the coefficient of the dependent variable, the use of BCT on the right-hand side variables of (6-A) with (6-D) is less directly linked to automatic error variance manipulation than the use of BCT on its left-hand side variable.

Logit core regression. The same anticipation of weak linkages between the form of explanatory variables and implied hypothetical corrections for heteroskedasticity holds in Logit formulation (6-B) with (6-E) where BCT are used only on right-hand side variables, a generalization that might as well have been called after Box & Tidwell (1962), who do not transform *y*, rather than after Box & Cox (1964), who do.

If the modal *choice probabilities* were each transformed by BCT in (4), the situation would become comparable to that of (6-A). This occurs indirectly when an *inverse transformation* is used on the $[\exp(V_i)]$ quantities of the Logit, for instance in Multinomial Linear Inverse Power Transformation-Logit (LIN-IPT-L) core expressions inspired by (6-C) and documented further in Section 7 below:

$$(12\text{-A}) \qquad \left\{ \left[\phi_m \exp(V_m) + 1 \right]^{1/\phi_m} - \mu_m \right\}, \quad \phi_m \geq 0 \text{ and } \mu_m \leq 1, \qquad \text{[BTG applied to Logit Quantity]},$$

$$(12\text{-B}) \qquad \left\{ \left[\phi_m \exp(V_m) + 1 \right]^{1/\phi_m} \right\}, \qquad \phi_m \geq 0, \qquad \text{[BTG applied to Logit Quantity]}.$$

Because the transformations of choice probabilities that implicitly occur in the so-called Linear IPT-L are strictly[52] equivalent to the direct transformation of *y* in (6-A), they could justify similar expectations[53] pertaining to the effects of the indirect transformations on error variance. In the "Box-Tidwell Logit" formulation (6-B) however, anticipations are quite different: one expects weak linkages between changes in the form of explanatory variables and hypothetical corrections for

---

[48] This issue is related to, but distinct from, that of the transformation of the residuals to normality. Concerning this last point, Draper and Cox (1969) have shown that this transformation can be useful even in situations where no power transformation of a variable can produce normality exactly but Nelson and Granger (1979) contest this claim after testing 21 series and finding few cases where the distribution achieves normality.

[49] This goes well beyond assuming that heteroskedasticity is proportional to one variable, as in Seaks & Layson (1983).

[50] An urban transit and automobile trip demand application (Dagenais *et al.,* 1987) is the source of Figure 4.

[51] In classical regression, the coefficient of the dependent variable is constrained to 1.

[52] This is readily understood if one remembers that the *M* dependent variables of the Multinomial Logit can be seen as resulting simply from a normalization of each by their sum: normalization is simply a scale factor.

[53] The Pregibon (1980) binomial link used for aggregate (share) data, based on a direct Box-Cox transformation of the $[\exp(V_1)]$ quantity that includes the Logit link as a special case, also involves a transformation of the choice probability that could make it difficult to modify form without significantly affecting error variance.

heteroskedasticity. The tests carried out on the Swedish "Box-Tidwell Logit" model just discussed in Table 5 entirely confirm this anticipation, as we presently recall.

C. Form and heteroskedasticity in the Swedish models for business and private trips

A reasonable expectation of misspecification in that model as specified had to do with the difficulty of taking into account in the construction of explanatory factors such as overnight stopping costs related to trip length, the nature of which might have induced higher error variance.

It was consequently assumed, for each trip purpose p and observation n, that the correction factor $\mu_i[f_i(Z_i)_n]^{1/2}$ be equal to 1 if the distance covered by the trip was up to 300 km, equal to γ_{p2} if it was in the 301-600 km range, equal to γ_{p3} if it was in the 601-900 km range, and equal to γ_{p4} for trips beyond 900 km. This yielded 3 new parameters by trip purpose model, each *linked to distance but independent from the alternatives*. Of these 6 adjustment factors, only the γ_{p3} of business trips was found to be significantly lower than and distinct from 1; this significance also decreased the importance of segmentation but had no discernable effect on the BCT estimates or on elasticities. It was concluded that this result made sense but was nothing to write home about, as the reader might have anticipated all along from the above discussion on "Box-Tidwell Logit" specifications. This is in line with Schnetzler (1996, 1998), who convincingly argues that heteroskedasticity is generally secondary unless socio-economic factors really matter.

D. Form and autocorrelation

To the extent that form and heteroskedasticity are also linked to autocorrelation, it is natural to raise the issue of their joint determination and to explore the effect of autocorrelation in the determination of form.

It was explicitly raised in a time-series model of transit and automobile demand consisting in 3 equations of type (6-A) where due weight was given to multiple-order autocorrelation and to heteroskedasticity (6-D) considered jointly. The authors concluded (Dagenais *et al.,* 1987, p. 460): "The BCT can be shown to dominate either the linear or the log form unconditionally, *i.e.* independently from whether one simultaneously takes into account [various forms of] heteroskedasticity and [of multiple-order] autocorrelation."

In Logit models of type (6-B), the addition of (1st order and 12th order) autocorrelation to an aggregate[54] binomial transit trip mode of payment model (Gaudry & Wills, 1979) made no real difference to the optimal BCT estimates, found in that case to lie numerically close to the logarithmic case but to differ statistically very little from the linear values.

We could not find examples of applications with discrete models where serial (or more general, for instance spatial) autocorrelation were estimated simultaneously with form parameters, despite the taking into account of serial autocorrelation since about 1985, as documented in surveys (Azzalini, 1994; Jackman, 1998), notably in Mixed Logit models where the specification of residuals is becoming extremely sophisticated (Ortúzar, 2006).

One or more of these model dimensions might smell of diminishing returns to complexity, but which one? In particular, how much does it matter in practice that BCT values be this or that and trade-offs among modal characteristics[55] not be constant? More generally in fact, what values should BCT parameters obtain and can they be made sense of?

---

[54] It is of course known that market share models will do well in time-series estimation in the presence of population heterogeneity (Givon & Horsky, 1978).

[55] In the forthcoming four Tables 7-11, we deal only in BCT cases. Binomial urban cases have been explored with cubic spline (Gillen & Cox, 1979) and simple power (Koppelman, 1981) functions; a meticulous Air-HSR example has been developed with polynomials (Blayac & Causse, 2001).

5. Forms interpreted: "Cost damping" and attitudes to Risk or Distance

The cost damping query. The subject of anticipated BCT form values can be conveniently addressed as an answer to the first question recently asked by the United Kingdom Department for Transport about the existence of absolute Cost damping, defined as "a feature in some models that the impact of Cost and/or Time is reduced for longer journeys" (Daly, 2008, 2010).

A second question arises about the existence of relative damping of Cost and Time impacts based on the empirical observation that "Value-of-time studies in the transport sector most often find that the Value of travel Time in money terms increases with the length of the trip" (Daly, 2010, p. 5), as confirmed in the latest meta-analysis (Abrantes & Wardman, 2011, Table 15).

Cost damping and form in step. We relate the damping claims to BCT form values in five steps corresponding to sub-sections:

i) First, we define the required Demand sensitivity measures in question and show that the existence of negatively sloped curves, caused by diminishing marginal utility, is analytically independent from that of BCT applied to model variables, as they can in principle take any value without affecting the key negative sign property of demand slopes for normal goods.

However, actual form parameter estimates determine whether in fact damping occurs or amplification, its opposite, prevails. Damping or amplification occurs for some defined value ranges of the relevant BCT, domains separated by a border where independence holds.

But, as Demand slope sensitivity depends on the form of many transport variables, we start our examination of value ranges with LOS variables, distinguishing between absolute and relative damping and drawing from three tables containing survey results.

ii) Second, turning to actual BCT estimates for Time and Cost variables of intercity passenger models in Table 7 (and of freight models in Table 9), we find that they are generally consistent with the existence of *absolute* Time and Cost *damping*. By contrast, in urban passenger models found in Table 8, Time *amplification* typically occurs concurrently with Cost damping. This finding establishes the existence of a fundamental difference between intercity and urban travel.

iii) Third, in a similar analysis of rates of substitution between Time and Cost variables of all models, *relative* damping occurs in all tables (7-8-9) with a couple of exceptions. The frequency of damping/amplification can be plotted (Figure 5) or categorized (Figure 7) to focus on the specifics of the small minority of amplification cases.

iv) Fourth, overall, we find the systematic flexibility afforded by BCT to be well adapted to answering the questions raised on LOS variable sensitivity. Interestingly, the absence of absolute and relative sensitivity in popular structures based on *log-sum* dependent Total demand models and Linear Logit mode choice models requires that critical BCT borderline values between damping and amplification hold. Such values imply a very special behaviour of the Transport decomposition (Figure 6) of a movement along a demand curve and, we suspect, of traditional microeconomic decompositions between Income and Substitution effects as well.

v) Fifth, to gain some insight into atypical (absolute and relative) amplification cases identified by BCT values, we use in Table 10 new approaches, focusing on Logit models enriched by the multiplication of LOS factors (raised to a BCT power) either by a Risk attitude term (using a simple power) defined under Rank Dependent Utility (RDU) postulates or by a Distance attitude term (also raised to a simple power).

The latter interaction allows for a clear distinction to be made in gross BCT damping and amplification power results between an element measuring "attitude to Distance" and another "attitude to Outcome" proper, giving rise to a "calculus of gross and net effects" of Distance.

5.1. Form parameters and Modal demand slopes under Certainty (UC)

For any demand function, sensitivity with respect to Price or Time can be expressed as a partial derivative or slope (with units), or as an elasticity (without units). Considering the full Demand formulation (1)-(2) applied in (7-A), we state the better known elasticity measure first but proceed to discuss sensitivity with the slope measure. Answering the query requires defining damping and amplification, notions that also involve second derivatives. We now assume that LOS are certain.

The classical elasticity of the Demand for a particular mode, such as rail, with respect for instance to its Fare[56], is expressed in a QDF framework in the familiar way (*e.g.* Oum *et al.*, 1990, p.7) as the sum of elasticities for total traffic and mode split, for instance with a Box-Cox Logit core:

$$(13\text{-A}) \qquad \frac{\partial T_{rail}}{\partial X_{rail,Fare}} \frac{X_{rail,Fare}}{T_{rail}} = \left\{ \beta_U \frac{U^{\lambda_U}}{T_{TOT}^{\lambda_{TOT}}} \bullet \beta_{rail,X_{Fare}} X_{rail,F}^{\lambda_{rail,X_F}} P_{rail} \right\} + \left\{ \beta_{rail,X_{Fare}} X_{rail,F}^{\lambda_{rail,X_F}} (1 - P_{rail}) \right\}$$

where, to be precise[57], the last right-hand side (RHS) expression is the elasticity of the rail Mode share (or choice probability) P_{rail} with respect to the rail Fare, and the first RHS one is the elasticity of the Total demand T_{TOT} with respect to U multiplied by elasticity of U with respect to that Fare.

To study the sensitivity of rail demand, we first isolate the partial derivative and collect terms:

$$\frac{\partial T_{rail}}{\partial X_{rail,Fare}} = \left\{ \beta_U \frac{U^{\lambda_U}}{T_{TOT}^{\lambda_{TOT}}} \bullet \beta_{rail,X_{Fare}} X_{rail,F}^{(\lambda_{rail,X_F}-1)} P_{rail} \bullet T_{rail} \right\} + \left\{ \beta_{rail,X_{Fare}} X_{rail,F}^{(\lambda_{rail,X_F}-1)} (1 - P_{rail}) \bullet T_{rail} \right\}$$

$$(13\text{-B}) \qquad \frac{\partial T_{rail}}{\partial X_{rail,Fare}} = \beta_{rail,X_{Fare}} X_{rail,F}^{(\lambda_{rail,X_F}-1)} T_{rail} \left[\beta_U \frac{U^{\lambda_U}}{T_{TOT}^{\lambda_{TOT}}} P_{rail} + (1 - P_{rail}) \right]$$

where β_U, associated with the coupling term U, is assumed to be positive and, all other terms being necessarily positive except for β_{rail} which duly gives the negative slope sign, *Fare sensitivity* depends on three variables affected by corresponding BCT exponents [λ_F, λ_U, λ_{TOT}]. Characterization of damping and amplification combines first derivative (13-B) with results from Table 1 or from the second derivative $\partial^2 T_{rail} / \partial (X_{m,Fare})^2$, where $X_{m,X}$ denotes the rail Fare:

$(13\text{-C}) \qquad \dfrac{\partial^2 T_m}{\partial (X_{m,X})^2} = \beta_{m,X} X_{m,X}^{(\lambda_{m,X}-2)} \left\{ (\lambda_{m,X} - 1) \beta_U \lambda_U \dfrac{U^{\lambda_U}}{T_{TOT}^{\lambda_{TOT}}} P_m + (1 - P_m) + \left[\beta_{m,X} \lambda_U U^{\lambda_U-1} e^{V_m} X_{m,F}^{\lambda_{m,X}} \right] \right\} T_m$

namely, considering without loss of generality $\lambda_{TOT} = 0$ and the *log sum* case $\lambda_U = 0$:

$$\partial^2 T_m / \partial (X_{m,X})^2 = \beta_{m,X} X_{m,X}^{(\lambda_{m,X}-2)} \left\{ (1 - P_m) \right\} T_m$$

which, for the domains of relevance for our purposes, has the property that, if

$\lambda = -1$ then $\partial^2 T_m / \partial (X_{m,X})^2 = \beta_{m,X} / X_{m,X}^3 \left\{ (1 - P_m) \right\} T_m$,   *i.e.* decreases with $X_{m,X}$

$\lambda = 0$ then $\partial^2 T_m / \partial (X_{m,X})^2 = \beta_{m,X} / X_{m,X}^2 \left\{ (1 - P_m) \right\} T_m$,   *i.e.* decreases with $X_{m,X}$

$\lambda = 1$ then $\partial T_m / \partial X_{m,X}$ in (13-B) is independent from $X_{m,X}$ and $\partial^2 T_m / \partial (X_{m,X})^2$ does not exist

$\lambda = 0,5$ then $\partial^2 T_m / \partial (X_{m,X})^2 = \beta_{m,X} / X_{m,X}^{1,5} \left\{ (1 - P_m) \right\} T_m$,   *i.e.* increases with $X_{m,X}$

$\lambda = 2$ then $\partial^2 T_m / \partial (X_{m,X})^2 = \beta_{m,X} \left\{ (1 - P_m) \right\} T_m$,   *i.e.* is an inflexion point, constant with $X_{m,X}$

$\lambda = 3$ then $\partial^2 T_m / \partial (X_{m,X})^2 = \beta_{m,X} X_{m,X} \left\{ (1 - P_m) \right\} T_m$,   *i.e.* increases with $X_{m,X}$

---

[56] We assume that the LOS variable appears only in own-mode representative utility function and will not develop the symmetrical expression for *Time sensitivity*. For an analysis of conditions under which choice and demand elasticities can be identical, see Smith *et al.* (2010).

[57] Of all possible cases considered at the end of Appendix B, the combination (43-G)-(43-H) suffices at this point.

The Fare exponent (λ_F - 1), and by symmetry that of Time (λ_T - 1) if the demand slope is formulated with respect to Time, as well as the exponent of Utility λ_U and the exponent of total market size λ_{TOT}, have no effect on the sign of slope (13-B). But they will modulate it, *i.e.* make it sensitive to the level of Cost (or Time) and to that of Utility U or total market size T_{TOT}. BCT values can increase, decrease, or have no effect on, the Demand slope.

Because the "Cost damping" query is formulated in terms of Time and Cost factors, we start the discussion of the role of BCT power exponents of the four candidate variables with the level of service (LOS) variables, momentarily neglecting the Utility and Total Market size (UT) variables.

5.2. LOS form parameter values and absolute damping in three market types

Starting with the value of their exponents, found in Columns 2 and 3 of Table 7 for intercity trips, the LOS sensitivity of the slope will be said to exhibit:

		Effect of ΔX on slope curvature			
(13-D)	absolute amplification	╲	→	╴	if (λ_{LOS} - 1) > 0,	*i.e.* when λ_{LOS} > 1
(13-E)	absolute independence	╲	→	╲	if (λ_{LOS} - 1) = 0,	*i.e.* when λ_{LOS} = 1
(13-F)	absolute damping	╲	→	╲	if (λ_{LOS} - 1) < 0,	*i.e.* when λ_{LOS} < 1

which simply means that "Cost damping" as understood in the query above is defined in (13-F) as a slope that increases in absolute value, but at a decreasing rate. Also, following the same logic, "Cost amplification" is defined in (13-D) as a slope that increases in absolute value, but at an increasing rate. Naturally, independence is the borderline case (13-E) of a slope that is constant with LOS.

We observe in Models 1-7 and 10-16 of Table 7[58], where the Logit utility functions are defined as in (6-B), that *absolute damping* (λ_F or λ_T smaller than 1) occurs in all 28 possible values but one, found in Model 2, where the Time exponent value is 1,80. Absolute Fare and Time damping are therefore both pervasive. This is consistent with some analysts casually "finding that the logarithm of variables provides better estimation results without necessarily relating theses results to either perception transformations or to marginal utility (Quarmby, 1967; Stopher & Lavender, 1972)", as pointed out long ago by Koppelman (1981).

But the overall prevalence of absolute damping (13-F) over amplification (13-D) is not limited to intercity passenger models of Table 7, as the urban examples listed in Table 8[59] and the freight examples in Table 9[60] indicate: three quarters of the cases exhibit LOS damping in these two new tables, with the exceptions found in urban markets where Time amplification *always* occurs, as can be seen by the braided boxes in Column 2 of Table 8. This is consistent with, and generalizes what may be called, the "urban amplification" finding of Levin *et al.* (1980) "that car drivers and bus riders overestimate travel time and specifically that such overestimates increase with trip duration", as again summarized by Koppelman (1981, p. 131).

---

[58] Table 7 does not include cases estimated with very coarse grids and by hand even when they appear to display both absolute and relative damping for the modes considered, such as Car and Public transport in Vtric *et al*. (2007).

[59] Table 8 does not include the study on the temporal stability of discrete choice models by McCarthy (1982) where the functions estimated with BART data appeared linear whether one used two modes (Car and Bus, before BART) or a more complex break-down of the public mode into 3 sub-categories (after BART). This finding remains an exception and we could not determine from the paper whether peculiarities of local pricing (such as bus Fare varying over a very narrow domain) could explain the result.

[60] Table 9 does not include the model used by BVU (Schneider, 2003; Selz, 2004) to produce freight forecasts for the German Federal Transport Infrastructure Plan (FTIP) of 2003 because BVU, despite the public nature of the German FTIP, refused to supply information on the values of the two BCT used for Time and Cost variables. Neither was it possible to find out if BVU, as stated in Selz's presentation (2004, page 20), is really using Modal utility functions without intercepts, in which case their BCT estimates would not be invariant to units of measurement of the variables (Schlesselman, 1971). In some past work with intercity passenger Logit models (Kessel *et al*, 1986), BVU did not use any modal constants.

Table 7. BCT estimates for Induction, Time & Cost variables in intercity passenger models

Models*	Column	1	2	3	4	
National		**Utility**	**Logit expense specification**			**Source in this paper: results obtained from**
(Domestic only, except 9)	**Purpose**	λ_U	λ_{Time}	λ_{Fare}	$(\lambda_{Time} - \lambda_{Fare})$	
1. Canada 1972 (4 modes)	All	0,05	-0,05	-0,26	0,21	Table 2
2.-3. Quebec-Windsor Corridor Canada 1987 (4 modes)	Business	---	1,80	0,28	1,52	Table 3, Col. 2; see (1)
	Other	---	-0,12	-0,29	0,17	Table 3, Col. 5; see (1)
4.-5. Quebec-Windsor Corridor Canada 1991 (4 modes)	Business	---	0,56	0,25	0,21	Table 4, Col. 2
	Other	---	-0,10	0,31	-0,41	Table 4, Col. 4
6. Sweden 1984-85 (4 modes)	Other	---	0,15	0,40	-0,25	Figure 3 model
7. Germany 1979-80 (3 modes)	All	---	*0,24*	*0,24*	*0,00*	Figure 10 model
		Utility	**Logit rate specification**			**Source**
	Purpose	λ_U	λ_{Speed}	λ_{Price}	$\lambda_{Distance}$	**Gaudry *et al.*, 1994**
8. Canada 1976 (4 modes)	All	-0,08	0,15	1,63	-0,25	T.B.1, C3; T.B.2, C.6
9. Germany 1985 (3 modes)	All	0,41	6,39	-2,20	-0,15	T.3, C.6; T. 2, C.3;
Multinational (3 modes)		**Utility**	**Logit expense specification**			
(Domestic and cross-border)	**Purpose**	λ_U	λ_{Time}	λ_{Fare}	$(\lambda_{Time} - \lambda_{Fare})$	**Mandel, 1992**
10. Germany 1979-80 (3 m.)	All	---	−0,56	−1,09	0,53	Tableau 3.2.1; see (2)
						Gaudry *et al.*, 1998
11.-13. France 1993-1994, Cross-Channel: see (3)	Business	0,33	*-0,62*	*-0,62*	*0,00*	App.2, C.5; App.1, C.1
	Vacation	0,12	*0,53*	*0,53*	*0,00*	App.4, C.5; App.1, C.6
	Private	0,31	*-0,33*	*-0,33*	*0,00*	App.3, C.5; App.1, C.4
						Mandel, 1999
14.-16. Surveys (see 4) in all international German airports (circa 1991)	Business	---	*0,40*	*0,40*	*0,00*	Table 2; or Table 3-10 in **Gillen *et al.,* 2001**
	Vacation	---	0,64	0,01	0,63	
	Private	---	1,00	0,70	0,30	

* All use discrete mode choice data, except for 1, 8 and 9 which are based on aggregate mode share data, and Revealed Preference (RP) data, except 4 and 5 which are based on face-to-face Stated Preference (SP) data. None use computer assisted telephone interviews (CATI).

(1) If one adds time by car to all Modal utility functions, as done in Column 3 and Column 6 of Table 3, the resulting Generalized Box-Cox specification differs from the Standard Box-Cox Logit specification used in all other models: we therefore choose Column 2 and Column 5 cases.

(2) The KONTIFERN data for Germany and abroad, the same as used in Model 10, were aggregated into modal shares to use an available algorithm which also allowed for Dogit and Inverse Power Transformation-Logit (see Equation 12) specifications applied to shares.

(3) Includes United Kingdom International Passenger Surveys (1991 and 1996) and Civil Aviation Authority (CAA) data sets and about 13 500 trips from Germany (Mobility'95), as well as trips for Norway and Sweden. The mode choice models are estimated with 77 568 observations (business, 22 579; private, 17 477; vacation, 37 512). The parameters were used for the first version of VACLAV-VIA (Schoch, 2000, 2003). A subset of relations within the Czech Republic pertains to 4 modes.

(4) About 238 000 trips. The data were checked by D. L. R. (Cologne). Equality of the BCT was imposed for business trips because the increase in Log Likelihood with distinct Time and Cost was slightly less than 2,33.

5.3. LOS form parameter values and relative damping in three market types

But, should there instead be restrictions or strong expectations based on marginal rates of substitution among modal characteristics such as Time and Fare? The value of time (VOT) obtained by dividing Time and Fare derivatives, each obtained in turn from (13-B), is simply:

$$
(14\text{-}A) \qquad VOT \equiv \frac{\partial T_{rail}/\partial X_{rail,\,Time}}{\partial T_{rail}/\partial X_{rail,\,Fare}} = \frac{\beta_{rail,\,X_{Time}}\, X_{rail,\,Time}^{(\lambda_{rail,\,X_{Time}}-1)}}{\beta_{rail,\,X_{Fare}}\, X_{rail,\,Fare}^{(\lambda_{rail,\,X_{Fare}}-1)}}
$$

where, again, Time and Fare variables are always positive independently from the values or signs of their exponents, respectively (λ_T - 1) and (λ_F - 1). It seems that theory puts no more particular constraints on the ratios of BCT than on the levels of BCT for any particular variable[61].

---

[61] The flexibility of the BCT form avoids the fixed form dilemmas of old where the marginal rate of substitution had to be estimated either with fixed ratios of Time and Cost regression coefficients or with constant values of these coefficients provided by a given functional form (*e.g.* Gronau, 1970, Ch. 5).

What can then be made of the important second part of the query, relating sensitivities to trip length or Distance? To address it, we distinguish between two ways of giving a role to Distance and consider the following competing general Expense and Rate specifications of modal characteristics in representative utility functions:

(15-A) f_{iE} (Fare$_i$, Time$_i$),   [Expense specification]

(15-B) f_{iR} (Price$_i$, Speed$_i$, Distance$_i$),   [Rate specification]

where the money Price per unit of distance is obtained by dividing the Origin-Destination (OD) Fare by Distance from origin to destination and the time price per unit of distance, Speed, results from a similar operation applied to OD Travel time.

An explicit choice between Expense and Rate specifications. If the optimal form of the utility function is logarithmic, (15-A) and (15-B) are indistinguishable and yield identical log-likelihood values; otherwise, they are distinct but not nested. When BCT are used, it is very often the case that the Rate specification dominates the Expenditure specification both in terms of log-likelihood (applying the mechanics of non nested comparisons[62]) and in terms of muticollinearity. In that sense, the choice between an Expense and a Rate specification is the first to be made as one specifies representative utility functions. But if an Expense specification has been estimated, it is always possible to rewrite the explicit Expense estimates in implicit Marshallian Rate metrics.

The Expenditure specification, normal in Logit models, is still the most frequent in practice, but the Rate specification is closer to micro-economic basics where the size of the purchased basket depends on Income and its composition depends on Prices: in (15-B), Distance seems to play a role similar to that of Income, with money and time unit prices determining basket mix. In that sense, it functions as a "Time-income" proxy (Dagenais & Gaudry, 1986) driving trip lengths while the modal characteristics modify modal mix. But we shall see below that it may have another meaning.

Bringing out the latent role of distance in Expense specifications. Some insight into the damping issue can be gained from isolating from estimated parameters of (14-A) the implicit effect of Distance, a rewriting that was not necessary to define absolute damping above. It simply decomposes the Fare as a product of Price and Distance and the Time as a product of the inverse of Speed and Distance:

(14-B)
$$VOT = \frac{\beta_{rail, X_{Time}} \left[V_{rail, Speed}^{-1} \right]^{\lambda_{rail, X_{Time}} - 1}}{\beta_{rail, X_{Fare}} P_{rail, Price}^{\lambda_{rail, X_{Fare}} - 1}} D_{rail}^{\lambda_{rail, X_{Time}} - \lambda_{rail, X_{Fare}}}$$

where the estimated parameters from (14-A) are left unchanged by the rewriting, VOT sensitivity in (14-B) is a function of the arithmetic difference between the originally estimated Time and Fare powers ($\lambda_T - \lambda_F$) assigned to Distance, and can be said to exhibit:

		Effect of ΔX on Value of Time	
(15-C)	relative damping	▲	if $(\lambda_T - \lambda_F) > 0$,
(15-D)	relative independence	=	if $(\lambda_T - \lambda_F) = 0$,
(15-E)	relative amplification	▼	if $(\lambda_T - \lambda_F) < 0$,

which simply means that "relative cost damping" as understood *in the query[63] above* is defined in (15-C) as a VOT that increases with Distance. Also, by the same logic, "relative cost amplification" is defined in (15-E) as a VOT that decreases with distance. Naturally, relative cost independence is the borderline case (15-D) of a VOT that is constant with Distance.

---

[62] Basically building, at least conceptually, an artificial model that includes the variables from both specifications and afterwards comparing the nested specifications (15-A) or (15-B) to the artificial one formed with their union.

[63] It may seem bizarre to call damped a ratio that increases. One could reverse labels (15-C) and (15-D) and apply the damping notion to the ratio itself but that would modify the wording of the UK Department for transport query.

To find out if relative cost damping occurs in practice, we presently consider ($\lambda_T - \lambda_F$) estimates not only in intercity passenger models but also in urban passenger and in freight models. It will be clear that Daly's observation — that the VOT measured by (14-A) typically increases with trip length — holds[64] in the great majority of cases.

A. Relative cost damping in long-distance passenger markets

In Column 4 of Table 7, the exceptions to the presence of relative cost damping are for non-business trips in the Stated Preference survey made in the Quebec-Windsor Corridor in 1991 (Model 5) and for the Swedish long-distance travel (Model 6). For Germany, where trips are the shortest among the 4 sampled regions, Model 7 with 1979-1980 discrete RP data (to be used below in Figure 10) yields a null difference because there was too small a statistical gain to relaxation of the common BCT for Time and Fare equal to 0,24, *i.e.* equal to the fourth root of these variables.

In that case, the implicit VOT calculated from (14-A) does not vary with trip length but still varies with the reference levels of the variables and with the size of increments considered from a given reference Fare or Time: absolute damping occurs, but relative damping cannot occur. However, when the *same* data are aggregated into shares in Model 10, both Time and Cost BCT are significant (a third BCT on frequency, not shown in Table 7, is jointly estimated at 0,99) and relative damping in fact obtains.

The explicit role of distance in Rate specifications. What happens if Distance is used as an explicit regressor within a model directly formulated in accordance with the Rate specification?

To answer this question, we use a comparative study of Canadian and German intercity passenger markets not mentioned above (Gaudry *et al.*, 1994) where the specifications of the national models were almost identical within a QDF-type framework built with aggregate 1976 data for Canada and 1985 data for Germany. In the Generation-distribution[65] part of the framework summarized in Part I of Table 22, the models of type (6-A) took heteroskedasticity and spatial or more general autocorrelation into account. In the mode choice part, from which the coupling term (3) arose, specification (15-B) was chosen after comparisons with (15-A) and, in both mode choice models, 4 BCT were estimated[66], including one for Distance and the others for Price, Speed and Frequency of service: results are summarized as Models 8 and 9 of Table 7.

In many samples, notably freight ones, modal OD distances D_m vary too little among the modes to be interestingly distinguished and the default common measure D used *per force* must be treated as a socio-economic variable; all modal coefficients of Distance β_{mk} can then be identified only if distinct BCT are used for each and all of the D variables inserted in the M utility functions: in the linear case, only differences with respect to a reference alternative can be estimated. However, in the Canada-Germany passenger model comparison, the D_m did vary enough within each country for generic regression coefficients to be estimated, with or without help from BCT. The coefficient of D_m was negative in both countries with a generic BCT equal to -0,15 for Germany and to -0,25 for Canada. Such comparable Canada-Germany results appear to confirm the interpretation of Distance as an "Income" term[67] reducing the levels of all modal utilities as trip length increases, but reducing them at a decreasing rate implied by the negative powers associated with D_m (case (13-F)).

---

[64] We do not discuss the standard errors of estimates: the rare slightly negative differences may not differ much from 0.

[65] For an urban Generation-Distribution example comparing (15-A) and (15-B) specifications for two modes, see Dagenais & Gaudry (1986).

[66] For reasons stated above concerning "Box-Tidwell" Logit models, heteroskedasticity tests were deemed unnecessary.

[67] Note that, for some long trips in Canada (*e.g.* Halifax to Victoria), going by plane instead of by land modes (a 6200 km road distance) can save a week of your life in each direction.

B. Relative cost damping in urban passenger and intercity freight markets

After noticing Time power values systematically above 1 in Table 8, it does not come as a surprise that the difference between estimated Time and Fare powers ($\lambda_T - \lambda_F$) shown in Column 4 is then in effect always positive (implying relative damping or a *VOT rising with distance*).

Table 8. BCT estimates for Time & Cost variables in discrete RP urban Logit passenger models

Column		1	2	3	4	Source
		Time and Cost terms; expense specification				Source
Sydney (2 modes)	Purpose	λ_{Tww}	λ_{Tveh}	λ_{Fare}	(λ_{Time}-λ_{Fare})	**Hensher & Johnson, 1981; see (2)**
CBD trips (car and train)		see (1)				
17. Northern suburbs (1971)	Work	1,000	0,50		0,00	Table 1, Col. 1 (μ_k = 001)
Washington, DC (2 modes)						**Koppelman, 1981**
18. City-wide (1968)	Work	2,57		0,56	2,01	Table 2, Col. 6
Paris region (6 modes)						**Gaudry, 1985**
19. City-wide (1976)	Work	1,000	0,50		0,00	Table 3
						Hivert et al., 1988
20. Orly airport origin (1986-1987)	Private	1,08	1,08	0,42	0,66	Model 5.2, p. 46
Paris region (2 modes)						**Lapparent, 2004**
21. City-wide (1997, 11 variables)	Work	1,19	1,19	-0,89	2,08	Table 4.8, p. 135
Santiago de Chile						**Pong, 1991; and Gaudry, 1994**
A-1. CBD corridors (9 modes)						
22. Las Condes & San Miguel	Work	**0,13**	**1,37**	**-0,56**	**1,93**	Series I-B-G; see (3)
B-1. City-wide 1991 (11 modes)						**Parra Granifo, 1995**
23. Peak AM trips 7:30-8:30	Work	0,32	1,000	0,82	0,18	Table 4, Col. 1; see (4)
24. Off-peak AM trips 10:00-12:00	Work	0,31	1,000	0,69	0,31	Table 4, Col. 2; see (4)
25. Peak AM trips 7:30-8:30	Study	0,21	1,000	-0,01	0,20	Table 4, Col. 3; see (4)
Time terms and [Cost/Income] ratio (see (5)) term; expense specification						**Pong, 1991, and Gaudry, 1994**
A-2. CBD corridors (9 modes)	Purpose	λ_{Tww}	λ_{Tveh}	$\lambda_{F/s}$	(λ_{Tveh}-$\lambda_{F/s}$)	
26. Las Condes & San Miguel	Work	**0,12**	**1,30**	**0,55**	**0,75**	Series I-A-G
						Gaudry et al., 1989
27. Las Condes (1983)	Work	0,44	1,56	0,23	1,33	Footnote 3 p. 156
28. Adding San Miguel (1985)	Work	0,33	1,57	0,60	0,97	Footnote 3 p. 156
B-2. City-wide 1991 (11 modes)						**Parra Granifo, 1995**
29. Peak AM trips 7:30-8:30	Private	0,46		0,53	-0,09	Table 4, Col. 5; see (6)
30. Off-peak AM trips 10:00-12:00	Private	0,54		0,64	-0,10	Table 4, Col. 6; see (6)
31. Off-peak AM trips 10:00-12:00	Study	1,00		0,25	0,75	Table 4, Col. 4; see (6)
Time terms and [Income - Cost] difference (see (7)) term; expense specification						
Paris region (2 modes)	Purpose	λ_{Tww}	λ_{Tveh}	$\lambda_{(I-F)}$	(λ_{Tveh}-$\lambda_{(I-F)}$)	**Lapparent et al., 2002**
32. City-wide (1997, 5 variables)	Work	1,17	1,17	-0,03	1,20	M-2 model; see (8)
						Lapparent, 2002
33. City-wide (1997, 5 variables)	Work	-0,05	1,11	0,07	1,18	M-2 model, p. 27;
						Lapparent, 2003
34. City-wide (1997, 16 variables)	Work	1,07	1,07	0,85	1,92	Table on page I; see (9)

(1) The value 1,000 denotes an untransformed variable appearing linearly in a model.
(2) In a previous analysis based on a single suburb subset (Hensher & Johnson, 1979b), the authors had found an optimal BCT value of 0,05 close to the logarithm but with a linear-probability model, not a Logit model.
(3) The income measure used is the net hourly wage rate.
(4) The Time variable denotes walk time.
(5) The Fare is divided by the net hourly Wage rate, in accordance with the Train-McFadden (1978) specification.
(6) The Time variable is a generalized time with weight of 1 for In-vehicle, 2 for Walk and 4 for Wait times.
(7) The Net Income term is obtained by subtracting Cost from Income.
(8) In Model 32, an equality constraint is imposed on the coefficients of total Time elements; it is relaxed in Model 33.
(9) In Model 34, 8 socio-economic dummy variables are added to the specification of Model 33. In consequence, the BCT on the Net Income variable becomes 0,85, *i.e.* almost linear and not significantly different from 1.

It is negative (implying VOT *falling with distance*) only in Models 28 and 29 where generalized time replaces and includes in-vehicle time. It might be that the low power values (in the range from

0,12 to 0,44) associated with walk and wait times when these variables are allowed specific BCT then dominate the In-vehicle time component of the generalized time construct, due to Parra Granifo (1995).

Table 8 also makes it possible to see how relative damping is affected by model specification:

i) *Cost to Income* specification*:* the table contains results of a thorough comparison between the specification of the goods/leisure trade-off as ratio of cost to income (Train-McFadden, 1978) in Model 26 and a more general specification where each term is self-standing in Model 22. The results from this unpublished work (Pong, 1991; Gaudry, 1994) on the high quality 1983-1985 Santiago de Chile data showed that the ratio specification 26 is easily rejected in favour of the unconstrained and more general specification 22 for all forms except linear ─ but that, among variants of 26, this linear case is completely dominated by free BCT form specifications that also happen to imply more reasonable VOT (not shown) than the single value imposed by the linear form. The Train-McFadden specification is preferred only if the choice is only between linear specifications of Model 22 and 26, where it is never the right form.

If the Train-McFadden ratio specification is strongly rejected, what then happens to relative cost damping? The comparison between results for the same data set (Las Condes & San Miguel), shown in bold for Models 22 and 26 (Series 1-A-G and 1-B-G), demonstrates that relative cost damping present in all listed urban models is here independent from whether one uses the rejected goods/leisure trade-off formulation or the freer specification more consistent with the data.

ii) *Net Income*: Table 8 also contains a comparison of the effect of adopting the Net Income formulation of the goods-leisure trade-off in Model 32 after having tried the usual specification in Model 21: the presence of relative damping is again unaffected by this change, as it is by freeing in Model 33 the constraint on the components of total time, or by increasing the number of socio-economic dummy variables in Model 34. In the latter case, this addition of "market segments" represented by 8 additional socio-economic factors almost linearizes the BCT on Net Income, as it should, but has no effect on the existence of relative Cost damping.

Table 9 presents some results for freight, all showing the presence of relative Cost damping in Models 35-37 estimated with an expense specification of the utility functions. In Model 38, the BCT is applied to a Distance variable assigned only to container and rail (truck is the reference).

Table 9. BCT estimates for Time & Cost variables in intercity freight Logit models

Column	0	1	2	3	4	
Discrete SP data	Lateness, Time and Cost terms; expense specification					**Source**
Norway (2 trucking modes)	λ_{Late}	λ_{Freq}	λ_{Tveh}	λ_{Fare}	$(\lambda_{Time}-\lambda_{Fare})$	**Fridstrøm & Madslien,**
Own account or for hire (1992)	See (1)					**2002**
35. All commodities	0,536	---	0,198	0,005	0,193	Table 1; see (2)
France-Spain border (2 modes)						**JLR Conseil & Stratec,**
36. Bulk commodities (2005)	1,000	1,000	0,700	0,100	0,600	**2005, p. 48**
37. Manufactured goods (2005)	1,000	1,000	0,800	0,400	0,400	Table p. 48; see (3)
Aggregate RP data			Rate specification			
France-Spain border (3 modes)			λ_{Speed}	λ_{Price}	$\lambda_{Distance}$	**Gaudry et al., 2008**
38. All commodities (1999)	---	---	1,000	-1,83	6,49	Annex 3, Table 11; see (4)
(1) The value 1,000 denotes an untransformed variable appearing linearly in a model.						
(2) The late delivery risk variable had a positive sign in the linear model specification. In the optimal model, a statistically significant negative sign obtains and the BCT equals 0,536.						
(3) Also found in Blardone (2007).						
(4) The road is the reference mode: Distance therefore appears with a common BCT in the utility functions of the two remaining modes: combined (container) road-rail and traditional rail.						

Damping and amplification frequency. Collecting information from Tables 7-9 makes it possible to define cross-fields in Figure 5 where higher frequencies drawn from the sample of all split

models (intercity passenger and freight, urban passenger) sharing the Expense specification can be represented by modulating the darkness of areas. We find that relative and absolute damping dominate in all market types except for urban areas where the value of λ_T systematically exceeds 1 and *always* implies absolute amplification.

Figure 5. Main concentration of BCT estimates for Time & Fare variables (*ex* Tables 7-9)
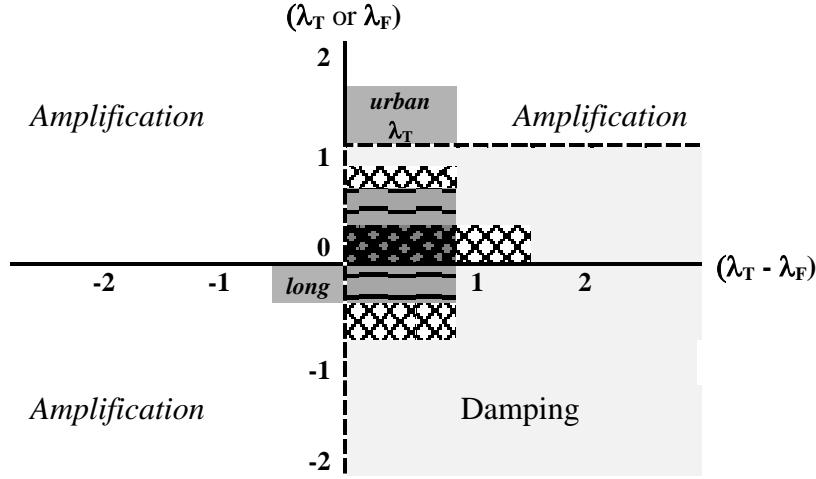
BCT flexibility and the cost damping query. All things considered up to this point, a belief in the existence of Cost damping, at least in the strict relative sense of the expression related to Distance but also in the absolute sense, is well served by the monotonic variable-specific flexible non linearity of the BCT: the belief requires demand function slopes (sensitivities) that are changing with the importance of LOS trip characteristics, and implicitly with distance, but specifically for each variable. In addition, BCT have facilitated our empirical discovery that Time amplification occurs in urban markets, in contrast with intercity markets, without affecting the general presence of relative damping. Appendix A implies that using simple powers of variables instead of BCT to achieve the same ends would not be so simple a matter.

5.4. Other "damping" factors and the decomposition of price effects

A. Completing the analysis of the influence of Time or Cost BCT on Demand sensitivity

We must now discuss the influence on the Demand curve slopes of Utility U^{λ_U} and Total market size $T_{TOT}^{\lambda_T}$, both UT terms momentarily neglected above to focus on the LOS terms. The influence of Total market size is of peripheral interest and we may set $\lambda_{TOT} = 0$, but the role of Utility remains central because it is the vehicle of induction. In that sense, Logit cores within a QDF are extensively defined by more or fewer restrictions on the form structure triplet [λ_U, λ_T, λ_F].

On the form of the Utility coupling term in intercity passenger models. In this respect, we note that actual estimates of λ_U found in Column 1 of Table 7 are concentrated between -0,08 and 0,41 and we remark in passing that values numerically close to zero, such as 0,05 in Model 1 and -0,08 in Model 8, are in fact significantly different from zero, as can be verified from the original studies[68]. But for the moment we concentrate on the five values of λ_U between 0 and 1.

In those five models, the sometimes blindly used *log-sum* obtained by imposition of the restriction $\lambda_U = 0$ is rejected in favour of positive values of λ_U that, in this range between 0 and 1, imply that LOS generate increased travel through U but at a decreasing rate, and certainly not at the constant rate implicit in the logarithmic case in (13-A) and (13-B). Does it matter?

---

[68] And, for Model 8, from Column 3 of Table 22.

We argue in Appendix D that it indeed matters critically in some trade models based simply on Distance; here we focus on the implications in the more sophisticated transport models specified with coupling terms built from the denominator of choice models. For this, we seek to understand some implications of restrictions on values of the triplet $[\lambda_U, \lambda_T, \lambda_F]$.

B. Form structures and the behaviour of price-quantity decompositions

To document the implications of the most frequently used among the different specifications of the triplet, namely $[\lambda_U = 0; \lambda_T = \lambda_F = 1]$, we first provide a graphical representation of the Transport decomposition of the effect of a Price change on Quantity demanded effected by introducing the transport quantity decomposition alone on a map of quantities, but without indifference curves.

In a second[69] step taken in Appendix B, this map is enriched by the superposition of the classical microeconomic decompositions due to Hicks (1939) and Slutsky (1915), which both require indifference curves, in order to show that all three decompositions could be studied further within the same graph if one wanted to illustrate how the implied pattern and shape of indifference curves changes under distinct form structure assumptions. But we will not study these map patterns, a complex enterprise in its own right, here or in Appendix B.

The Transport decomposition with two modes. Figure 6 shows how transport analysts often decompose, for any transport demand model[70], the change in the Quantity demanded subsequent to a lowering of a modal Price.

Figure 6. Transport decomposition of a variation in travel demand from 1 to 2
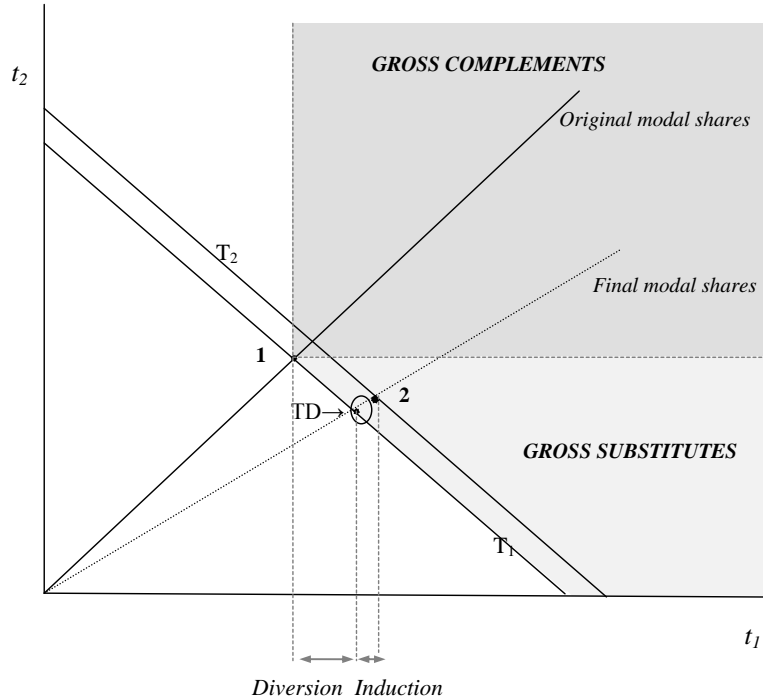
To understand the graph, first note that the two rays starting at the origin and passing through points 1 and 2 show the initial and final modal splits between t_1 and t_2 trips demanded for the two modes in question (modes 1 and 2); moreover, the intersection of any of these rays and of parallel total market lines T_1 or T_2 placed at 45° conserves the modal shares between modes 1 and 2. By contrast,

---

[69] The introduction is performed in many detailed steps in the source paper (Gaudry, 1998) centred on distinguishing substitution and complementarity features between HSR and air markets.

[70] This is easier to perform with QDF frameworks but can be effected with any framework: DDF, Armington, ...

any movement along these negatively sloped lines placed at 45° implies new splits between modal demands t_1 and t_2 summing to a constant Total number of trips.

Modal Diversion (or *transfer*) is the movement from point 1 to point TD where the modal shares have changed without affecting the total number of trips and *Modal Induction* (or *generation*) is the movement from point TD to point 2 where the total number of trips has changed without affecting the modal shares.

Figure 6 also identifies regions which, subsequent to a lowering of the price of mode 1, correspond to complementarity or substitution in classical microeconomics. Appendix B shows within the same graph (and under particular assumptions) identifiable differences between this Transport decomposition $\{[1 \rightarrow TD] ; [TD \rightarrow 2]\}$ and the known decompositions by Slutsky and Hicks.

Implications of different form structures. Consider the ratio of Induction to Diversion effects implied in the popular combination of a Linear Logit ($\lambda_T = \lambda_F = 1$) with a *log-sum* term ($\lambda_U = 0$). We can show that such imposed insensitivity to the level of LOS and U variables leads for the ratio to a simple expression, derived simply from components of (13-B):

$$(16) \qquad \frac{Induction}{Diversion} = \left[\frac{\beta_U \, U^{\lambda_U} \bullet \beta_{rail, X_{Fare}} X_{rail,F}^{\lambda_{rail, X_F}} P_{rail} \bullet T_{rail}}{\beta_{rail, X_{Fare}} X_{rail,F}^{\lambda_{rail, X_F}} (1 - P_{rail}) \bullet T_{rail}} \right] \Rightarrow \left[\beta_U \Bigg/ \left(\frac{1}{P_{rail}} - 1 \right) \right],$$

if we set $\lambda_{TOT} = 0$, which implies $U^{\lambda_U} = 1$. This Induction to Diversion ratio is indeed independent of LOS and Utility terms, and therefore constant, a property that might be surmised to have implications also for the ratios of Income to Substitution effects definable after Hicks and Slutsky to the extent that, for instance, constant marginal utility will imply for all three decompositions the very rigid and peculiar indifference map structure studied by Bruzelius (1979)[71].

The observed presence of damping or amplification in Tables 7, 8 and 9 above requires that Time or Cost slopes not be constant but influence the Induction to Diversion ratio driven by the shape of indifference curves. The *Log-Sum & Linear Logit* form structure [$\lambda_U = 0$; $\lambda_T = \lambda_F = 1$] disallows such systematic variability from the start[72] by requiring, according to Bruzelius, linearly additive utility functions demonstrated here to be untenable in most cases and samples considered.

Indeed, actual form parameter sizes and signs have in fact revealed more than movement away from a constant ratio such as (16). They have revealed *three systematic movements* and departures away from constancy: (i) non constant marginal utility of travel[73] (expressed in λ_U)[74]; (ii) a tendency of marginal utility of Cost and Time to diminish at a decreasing rate (expressed in λ_T and λ_F), with the exception of urban markets where utility of Time diminishes at an increasing rate; (iii) a tendency for the relative decrease with Distance of the marginal utility of Cost as compared to that of Time (expressed in the difference between λ_T and λ_F).

---

[71] See in particular his Appendix B of Chapter 3, pages 79-83, entitled: "Restrictions on the Utility function for the Marginal Value of time to be Constant".

[72] Some make a reasonable compromise: Setec International and Stratec (2008) set $\lambda_U = 0$ but estimate λ_T and λ_F for each of 5 intercity trip purposes in France. BCT values by trip purpose all apparently exhibit absolute and relative damping, as in most models of Table 7, but the single decimal place accuracy of the reported estimates makes it difficult to be highly certain of the pervasiveness of relative damping without access to all the digits actually used in the calculations. On such calculations, Spitzer (1982, Conclusion 4) recommends 4 decimal place accuracy.

[73] Remember that the comparison of Models 22 and 26 has led to a decisive rejection of the Train-McFadden goods-leisure trade-off.

[74] The rate of substitution between travel and other activities also naturally depends on λ_{TOT}.

It should be possible to give a visual representation of the implications of this demonstrated "non constancy" not only with the transport decomposition but with the two classical ones as well, and to contrast actual non constancy cases with hypothetical constancy cases, but we shy away from those complex graphic tasks and come back to Logit LOS cores, neglecting the Utility effects.

5.5. Logit cores and attitude to LOS uncertainty (UR) or to distance (DA)

We concentrated above on the establishment of non linearity, *i.e.* of variable marginal utility, as a real dimension of demand functions independent from the presence of non sphericity in distributions of model residuals and from consumer heterogeneity in any garb ("ordinary" segments or atomized "mixed" segments).

And our survey of non linear models found that the actual values of BCT powers associated with modal characteristics displayed obvious regularities discussed in the context of the "Cost damping" claims: for simplicity, we call those regularities under certainty assumptions the "Law of demand A" in summary Table 11. We now try to go further in two ways, identified in the same table as the "Law of Demand B" under added Risk attitude uncertainty assumptions and as the "Law of Demand C" under added Distance attitude postulates.

Seemingly atypical BCT estimates in all above models. Focus first in Figure 7 on a subset of results from Tables 7-9 to isolate amplification cases which seem atypical, limiting ourselves to the 21 Mode choice models in Expenditure format where *at least* two distinct BCT powers were estimated, a limitation that removes the noise from rough frequencies plotted in Figure 5. Generally speaking in Figure 7, absolute and relative[75] damping prevail and our concern is with a few notable exceptions: (a) the only intercity case of slope amplification pertains to business trips in the Quebec-Windsor Corridor (Model 2 in line B). All other cases of amplification pertain to urban areas; (b) the two cases of VOT amplification in line C, the first for the same corridor, but for non-business trips (Model 5), and the second for Sweden (Model 6).

Figure 7. Pinpointing amplification in BCT estimates for Time & Fare variables (*ex* Tables 7-9)

A summary of twenty-one			Power parameters of				Marginal rate of substitution	
Logit models where Time and Fare were separately transformed			TIME		FARE		VALUE OF TIME	
			$\lambda_T < 1$	$\lambda_T > 1$	$\lambda_F < 1$	$\lambda_F > 1$	$(\lambda_T - \lambda_F) > 0$	$(\lambda_T - \lambda_F) < 0$
• *Eight intercity passenger models*								
A	Models 1,3,4, 6,10,15		D		D		D	
B	Model 2 (business; RP data)			A	D		D	
C	Models 5 (SP) and 6 (RP)		D		D			A
• *Three intercity freight models*								
D	Models 35-37		D		D		D	
• *Ten urban passenger models*								
E	Models 18, 20-22, 26-28, 32-34			A	D		D	
Relative frequency of Damping and Amplification above and in Table 18								
Absolute Slopes	Damping	$\lambda < 1$	D*		D**			
	Amplification	$\lambda > 1$		A*				
Relative Slopes	Damping	$\Delta\lambda > 0$					D***	
	Amplification	$\Delta\lambda < 0$						A***
If Models 44, 47 and 48 of Table 18 are also considered, one finds in the 24 models (12 intercity and 12 urban):								
D*and A*	*Time damping occurs in 11/12 of the intercity models and in 1/12 of the urban models.*							
D*	*Fare damping occurs in 12/12 of the intercity models and in 11/12 of the urban models.*							
D* and A***	*Value of time is damped in 9/12 of the intercity models and 11/12 of the urban models.*							

---

[75] According to Jara-Diaz (2007, Equation 2.34, p. 61), VOT should always increase with Distance.

Introducing the attitude to LOS Risk and the attitude to Distance. To understand these special shaded amplification cases of Figure 7 in greater depth, we turn to recent theoretical developments made to take some account of LOS risk in the modelling of demand or to practical developments introducing an interaction between Distance and LOS terms.

What then about uncertainty in transport LOS conditions? Is it possible to identify and distinguish, within *prima facie* parameter values for Time or Fare, influences of *attitudes toward risk* from those pertaining to the *variable marginal perceptions* of these conditions? It is extremely difficult, if not impossible, to isolate risk attitudes from marginal utility valuations if the transformations of probability functions are linear; but we get a new shot at the problem with non linear functions[76].

Our first purpose is to understand how BCT power estimates for Time or Cost may be influenced by new implicit risk parameters γ_{Time} and γ_{Fare}, that, once recognized, cannot be added or subtracted from the former. The new parameters can be interpreted as revealing an optimistic, neutral or pessimistic "attitude towards Risk".

Our second purpose is to study another possible reason for atypical gross BCT values of LOS variables. In this case, where the new relevant power parameters isolating a new role for Distance (called an *attitude to Distance*) are again designated by the symbols γ_{Time} and γ_{Fare}, former gross BCT power estimates are algebraically increased or decreased by the presence of the newcomers in such a way as to allow for explicit calculations of gross and net values. They can also be interpreted as revealing an optimistic, neutral or pessimistic "attitude towards Distance". In particular, we develop the view that urban (and some intercity) trip markets may well display very different total (gross) BCT powers of Distance due to this attitude to Distance component γ distinct from the remaining outcome evaluation component λ. The distinction applies at least to Time and Cost but can also in principle be used for other door-to-door elements such as Walk time or Frequency of Service. Let us develop these attitudinal interactions in turn, starting with uncertainty.

Level-of-service variables under risk assumptions (LOS-UR). The central idea of Rank Dependent Utility specifications imagined by Kahneman & Tversky (1979), formulated by Quiggin (1979, 1982), discussed extensively in the aftermath (*e.g.* Chateauneuf & Cohen, 1994 or Cohen & Tallon, 2000) and generalized in Chateauneuf (1999), is that it is possible to distinguish, using a product of functions, between the attitude to outcome risk (attitude to risk or how the outcome probabilities are perceived) and the attitude towards the outcome itself.

In an empirical application of this idea for the first time in transportation, and to a Logit LOS Time variable, Lapparent (2004, 2010) chose as attitude to Outcome (AOU) function the BCT $\left(\lambda_{Time} \right)$. And, as attitude to Risk (ARI) transformation function $\psi : [0,1] \rightarrow [0,1]$ theoretically applied to a difference between two cumulative distributions of probabilities (p_1, ... , p_i, ... , p_K) ordered on a certain support involving a loss of resources (here In-vehicle Travel Time), namely

$$(17\text{-}A) \qquad ARI_i = \sum_{k=1}^{K} \left[\psi\left(\sum_{i=1}^{k} p_i \right) - \psi\left(\sum_{i=1}^{k-1} p_i \right) \right] \left[aou\left(Time_k \right) \right],$$

he chose the simple power function. His choice of a simple power function preserves the constraint that the sum[77] of perceived probabilities, called a capacity, equal 1, but other ARI functions

---

[76] Two curvatures now matter, but it is the perception of the distributions of probabilities of outcomes that is critically nonlinear here, not the attitude towards these outcomes which may exhibit linearity or not.

[77] This property does not hold for the BCT, often conveniently used as an Arrow-Pratt measure of constant relative risk aversion (concavity $\lambda < 1$ indicating in that context risk avoidance, and convexity $\lambda > 1$ a preference for risk),

respecting this constraint are available, such as the inverted S-shape function used by Tversky & Kahneman (1992) on the lines of Quiggin (1981), which embodies other assumptions[78] about the attitude towards risk, and even more complex constructs ─ see for instance the convenient tabulation in Stott (2006). The simple power function in ARI distorts the difference between the two cumulated values in (17-B): $\gamma < 1$ convexity contracts it, indicating optimism; $\gamma > 1$ concavity amplifies it, indicating pessimism; and a neutral pivot $\gamma = 1$ recovers the "untwisted" neutrality of the Expected utility of the outcome.

These choices jointly yielded the desired product formulation:

(17-B)	$$\sum_{k=1}^{K} \left[\left(\sum_{i=1}^{k} p_i \right)^{\gamma} - \left(\sum_{i=1}^{k-1} p_i \right)^{\gamma} \right] \left[T_k^{(\lambda_{Time})} \right]$$						
	allowing for a classification of co-monotonic cases with respect to time *loss*						
	$\gamma < 1$	$\gamma = 1$	$\gamma > 1$	$\lambda < 1$	$\lambda = 1$	$\lambda > 1$	Risk profile[79]
	Optimism	Neutrality	Pessimism	Dislike	Neutrality	Strong dislike	
i	●			●			Strongly risk seeking
ii	●				●		Weakly risk seeking
iii	●					●	Indeterminate
iv		●		●			Risk seeking
v		●			●		Expected value
vi		●				●	Risk averse
vii			●	●			Indeterminate
viii			●		●		Weakly averse
ix			●			●	Strongly averse
	Attitude to risk			**Attitude to outcome**			

Lapparent actually used 5 discrete Time frequency distributions varying with the period of year associated with two Air France paths between Paris Charles-de-Gaulle (Terminal 2) and two London airports (Heathrow and City) and assumed that professional frequent travellers belonging to his sample could be well aware of such distributions pertaining to the same morning time departure slot and airport; he did not transform the Fare variable.

He found, as shown in Model 39 of Table 10, that Time optimism ($\gamma_T = 0,54$) prevailed in ARI and that the most likely BCT value for AOU, ($\lambda_T = 0,98$), was slightly inferior to 1 but not significantly different[80] from 1, an interesting combination of risk optimism and neutral perception of time loss $[\gamma \neq 1, \lambda = 1]$, as in Yaari (1987), for this short[81] air hop.

---

including in Box-Cox Logit models (*e.g.* Montmarquette & Blais, 1987). The attitude to risk is reassigned to the $\psi(\cdot)$ function in the RDU framework of (17-B), the BCT now revealing only the attitude to outcome. The expression "aversion to risk" has become more ambiguous depending on whether a RDU $[\gamma, \lambda]$, or a simpler $[\gamma = 0, \lambda]$, framework is used.

[78] Decision-makers are assumed to be pessimistic when the probability of occurrence of a bad time outcome is low and optimistic otherwise, the inflexion point of the shape being driven by only one parameter: for instance, they refuse to participate in a lottery if the prize is low but not if it is high, as implied in Kahneman & Tversky (1979).

[79] This list of nested known cases and their risk profile labels were established with Lapparent. They correspond as follows: $[\gamma = 1, \lambda = 1]$, Case (v), to Bernouilli (1738, 1954); $[\gamma = 1, \lambda \neq 1]$, Cases (v) and (vi), to Neumann & Morgenstern (1947); $[\gamma \neq 1, \lambda = 1]$, Cases (ii) and (viii), to Yaari (1987); $[\gamma < 1, \lambda < 1]$ and $[\gamma > 1, \lambda > 1]$, Cases (i) and (iv), to Chew *et al.*, (1987); $[\gamma \neq 1, \lambda \neq 1]$, Cases (iii) and (vii), to Chateauneuf & Cohen (1994).

[80] This result obtained with a BCT is extremely close to that obtained for the same mode by Hjorth & Fosgereau (2011, Column 5 of Table 5 or 6) who unfortunately used a fixed form Te^{η_T} for the attitude to Time itself.

[81] As argued early on in this analysis, the smaller the sample domain of a variable, the harder it is to detect curvature and the less this matters if only marginal changes in LOS are considered.

Level-of-service variables under distance attitude assumptions (LOS-DA). Another way of enriching the problem at hand is to distinguish between the attitude to LOS and the attitude to Distance by lifting the restrictions by which they are *per force* linked in (14-B): one introduces an interaction between Distance and say Time, effectively setting the first component of (17-A) to 1 and enriching the $aou(Time)$ component, written more formally below[82] as $aou(x, g^n(x))$.

The idea of this third approach substituting for the attitude to Risk an attitude to Distance is that Distance, as revealed in everyday language, has a role in consumer utility functions independently from that of its implicit presence in LOS variables. Again, the mechanism will be one of a product of functions but we now return to the assumption that LOS service variables (the network) are provided under certainty and further assume that they are taken as given[83] by the traveller.

Table 10. Mode choice models distinguishing attitude to risk or distance from attitude to outcome

| | | Model | | Power Parameter Levels | | | | | | Power Parameter Differences | | | Mean | # |
| | | | | TIME | | | FARE | | | VALUE OF TIME | | | DISTANCE | of |
Meaning of γ parameter	Market	Modes	Purpose	γ_T	λ_T	Sum	γ_F	λ_F	Sum	γ_T - γ_F	λ_T - λ_F	Sum	in km	obs.
A. Denotes attitude to Risk														
39. Lapparent (2004), T. 5.1	Paris-London	2 air paths	Frequent	0,54	0,98	n.a.	0,00	1,00	n.a.	n.a.	-0,02	n.a.	343,00	756
B. Denotes attitude to Distance														
40. Lapparent *et al.* (2009), T. 6	Czech Rep.	4 modes	Occasional	-0,66	0,74	0,08	0,21	0,29	0,50	-0,87	0,45	-0,42	145,47	2044
41. Axhausen *et al.* (2008), T. 4	Switzerland	4 modes	All (mean)	-0,26	1,00	0,74	-0,60	1,00	0,40	0,34	0,00	0,34	42,89	15870
42. Ramjerdi (1993), T. 9.5.16	Norway	5 modes	Services	0,20	1,00	1,20	0,00	1,00	1,00	0,20	0,00	0,20	167,10	5824
43. Ramjerdi (1993), T. 9.5.28	Norway	5 modes	Work	0,40	1,00	1,40	0,00	1,00	1,00	0,40	0,00	0,40	180,79	5824
Attitude to Risk or Distance		Optimism:	γ < 1	Y			Y			Y			*Legend*	
		Pessimism:	γ > 1										*Assumed value*	
Attitude to Outcome	(Absolute	Damping:	λ < 1		Y	Y		Y	Y		Y			
	slopes)	Amplification:	λ > 1		Y			Y						
Value of Time	(Relative	Damping:	Sum > 0									Y	*Y = Amplification*	
	slopes)	Amplification:	Sum < 0									Y		
			Column	1	2	3	4	5	6	7	8	9		

Consider then a random variable x characterizing transport Run Distance, $x \in [\underline{D}, \overline{D}], \underline{D} > 0, \underline{D} < \overline{D}$, its associated distribution excluding the null value, $F : R_+ \setminus \{0\} \to [0,1]$, and a transformation function $\psi(\cdot)$ of that distribution with the same property as in (18-A), namely $\psi : [0,1] \to [0,1]$. Assume further that x is linked to an exogenously determined network Supply n drawn from a finite set of discrete possibilities $\mathrm{N}(x)$. The traveller may then be assumed to maximize utility on this predetermined distance interval offering a given fixed Supply:

$$(18\text{-A}) \qquad \int_{\underline{D}}^{\overline{D}} \max_{s \in \mathbb{N}(x)} \left[aou\left(x, g^s(x)\right) \right] d\psi\left(F(x)\right)$$

where, if the Supply actually consists in L service attributes, the interaction of attitude to Distance and attitude to attributes terms might be taken to be

$$(18\text{-B}) \qquad aou\left(x, g^s(x)\right) = \beta_0 + \sum_{los=1}^{los=L} \beta_{los} \left[h_{los}(x, \underline{D}, \overline{D}) \right] \left[aou_{los}(g_{los}^s(x)) \right]$$

---

[82] This formalization was generously contributed by Lapparent, prompted by a much less competent formulation in a previous draft.
[83] We therefore exclude from the Distance attitude formulation both LOS uncertainty and any potential endogeneity resulting, at least in aggregate models, from equilibration between demand and supply, but we do not exclude the possibility that Distance be observed with error, notably of the Berksonian kind (Berkson, 1950) where the true value X^{true} is distributed around observed values X^{obs}, so that ($X^{true} = X^{obs} + u$). In the classical case, the measured value is distributed around the true value and the observed value is equal to the true value plus an error ($X^{obs} = X^{true} + u$).

and where, maintaining $aou_{los}(g_{los}^s(x)) = T^{(\lambda_{Time})}$, the attitude to Outcome term defined in (17-B), obvious specifications for the attitude to Distance term certainly include

$$(18\text{-}C) \qquad h_{los}(x, \underline{D}, \overline{D}) = \begin{cases} \left(\dfrac{x - \underline{D}}{\overline{D}}\right)^{\gamma_{los}} \\ \dfrac{x^{\lambda_{los}} - 1}{\lambda_{los}} \end{cases}.$$

We choose the simple power candidate to match and account for current practice[84], say with Time:

$$(18\text{-}D) \qquad f(D, T) = \beta_T [D]^{\gamma_{D,T}} \left[T^{(\lambda_T)} \right],$$

where the Distance attitude power parameter has the same interpretation as that of the Risk attitude parameter in (17-B): $\gamma < 1$ convexity contracts objective Distance, indicating optimism; $\gamma > 1$ concavity amplifies it, indicating pessimism; and a neutral pivot $\gamma = 1$ is "untwisted" neutrality.

Beyond analogy, a feasible *pirouette*? In view of this, it is natural to ask whether, going beyond analogy of meanings for γ and the common use of products of functions, (17-B) could contain (18-D) as a special case, derived perhaps by making use of continuous functions and of suitable assumptions, and Distance itself be considered as an indicator of Risk. The exploration of this possible *pirouette* is beyond this survey but should be undertaken.

Distance attitude and trip length. The re-specification of Logit models with (18-D) interactions for both Time and Fare LOS variables leads to estimates of $\gamma_{D,T}$ and λ_T, to updated demand slopes of type (13-B) and we may consequently rewrite VOT ratio (14-B) under "Distance-attitude" assumptions as:

$$(18\text{-}E) \qquad VOT_{DA} = \frac{\beta_{rail, X_{Time}} \left[V_{rail, Speed}^{-1} \right]^{\lambda_{rail, X_{Time}} - 1}}{\beta_{rail, X_{Fare}} P_{rail, Price}^{\lambda_{rail, X_{Fare}} - 1}} D_{rail}^{\gamma_{rail, Distance Time} - \gamma_{rail, Distance Fare}} D_{rail}^{\lambda_{rail, X_{Time}} - \lambda_{rail, X_{Fare}}}$$

which means that *prima facie **gross BCT*** values obtained under uncertainty assumptions may in fact contain an explicit Distance attitude component that can be analytically "***netted out***" with (18-E), where[85] the new γ parameters effectively lift the previous constraints linking Expenditure and Rate metric exponents in (14-B).

This difference between gross and net values is calculated in Table 10.B[86] for Model 40 pertaining to the Czech Republic where both Time and Fare, treated as in (18-E), make **gross** VOT *amplification* appear (the sum of power terms equals -0,42) at the same time as **net** VOT *damping* obtains (the sum of BCT terms equals 0,45) after subtraction of the Distance attitude effect[87]. But the sample of available models in Table 10.B is insufficient to compute other meaningful net effects and reversals between gross and net amplification and damping.

Still, we note from the four available Models 40-43 cases that, despite the very restrictive AOU choices made in the last three, Distance power estimates all reveal an optimism ($\gamma_{Time} < 1$) that, as shown in Figure 8, tends to decrease with the average distance prevailing in the sample, but without

---

[84] We have found no application using a product of BCT applied concurrently to a Distance and to a LOS variable.

[85] The simplicity of this calculation results from the fact that interaction is formulated as a product of effects in (18-D).

[86] Table 10 does not include the two variants based on much smaller unusual computer assisted telephone interviews.

[87] In the preliminary version of their paper, Lapparent *et al.* (2009) indicate that the introduction of the interaction between Distance and other variables considerably increased the log-likelihood.

reaching pessimism proper ($\gamma_{Time} > 1$). All four sample average distances are short compared to those found in the Quebec-Windsor Corridor to be presently documented.

The longer, the worst in intercity markets. The tendency observed in Figure 8 is consistent with an "attitude to Distance" interpretation of Distance interaction models of structure (18-E) and may help to understand the high absolute amplification value $\lambda_T = 1,80$ found in that corridor for business trips (Model 2 of Table 7 or Figure 7) where the average distance[88] by the dominant mode, car (86,1% of all trips), is 337 km. It is possible that, in the often severe climatic conditions of that Canadian region, the underlying attitude to Distance reflects pessimism perhaps to the point that the pivot value $\gamma = 1$ is passed, and that the total result implies a high BCT value.

Relative amplification found for Sweden (Model 6) [89] and for non-business trips in the Corridor (Model 5 of Table 7 or Figure 7), where the car share is 90,0% and the average car distance covered is 300 km, may also be related to underlying distance despite the fact that, in all other models pertaining to Canada as a whole or to the Corridor (Models 1 to 4), and of course to Germany (Models 10, 15 and 16), relative damping prevails. This frequency (2 out of 9 cases) resembles that found in urban markets (2 out of 16 cases) where only Models 29 and 30 show amplification perhaps linked in those cases with a Time BCT not independent from the Frequency BCT.

Figure 8. Trip length increases Time Distance pessimism (Models 40-43)
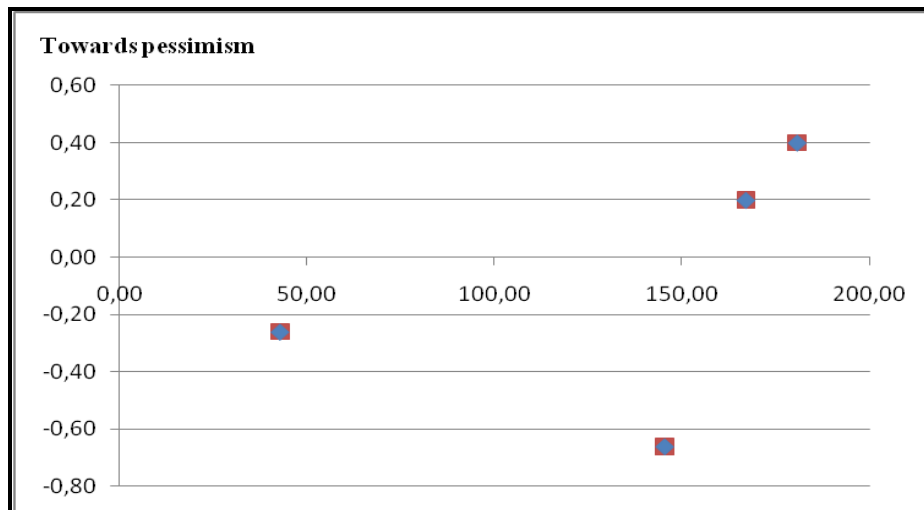
Pessimism in all urban markets. In urban models however, the extraordinary pervasiveness of Time amplification found in Table 8 and isolated in Figures 5 and 7 suggests the presence of relatively strong pessimism (perhaps even of positive values of γ_{Time}) arising from extensive daily experience[90] of urban networks: the gross linearity found in the BART case (McCarthy, 1982)

---

[88] When both car and rail are present, the average car distance is 356 km for business and 346 km for non business trips.
[89] The provisional results for the fourth milestone Swedish long distance model (Algers, 2011) based again, as for Model 6, on trips longer than 100 km, also exhibit relative amplification. This suggests that intercity VOT that decrease by Distance may be associated with cases of relatively long distances, like Canada and Sweden, as opposed to Germany or the United Kingdom. But only a study of attitude to Distance (LOS-DA) across countries differing markedly in trip length distribution could determine whether these are outliers and Daly's (*op. cit.*) second observation be amended.
[90] This interpretation seems consistent with a recent finding by Ben-Elia & Shiftan (2010). Inspired by Prospect theory, but in an experimental context different from that of Model 40, they found that experienced drivers took more risk in road path choices if they had good real-time information on the state of the network. As this information is notoriously bad in cities, their result would suggest that experienced urban drivers take little risk, or become quite pessimistic about time risk, which would mesh in with hidden positive values of γ_{Time} sustaining urban Time amplification values upwards.

commented in a footnote above might also be a case of relatively strong pessimism[91]. Suburban trains are not just slower High Speed trains: their Run Distances may be more resented.

Some other implications of the distinction between attitudes. This new interpretation of interactive Distance terms raised to a simple power has a side benefit in that, as each LOS variable such as Time is already accounting for Distance in model specifications under certainty, a further multiplication by Distance required a justification beyond improved fit. Since this *ad hoc* practice, used at least since Ramjerdi (1993)[92], was recommended in an extensive study of VOT (Mackie *et al.*, 2003), many authors (*e.g.* Hess, 2008; Axhausen *et al*, 2008; Lapparent *et al*., 2009) have, in models where utility functions contain both socio-economic variables (Distance and Income being considered as such by these authors) and network variables (at least Time and Cost), found goodness-of-fit gains to the addition of Distance interaction terms with such variables.

Are attitudes new? But it should not be thought that the conceptual distinction between attitude to risk and attitude to outcome, for one, is entirely new. As can be verified in summary Table 11, it was used by Marshall himself who, in his discussion of the postponement of consumption, believed in the objectivity of the perception of time uncertainty but in the subjectivity of trade-offs proper over time $[\gamma = 1$ for all individuals; λ changing across individuals$]$.

It is harder to find in the past explicit attitudes to Run Distance in the sense of (18-D) or, *mutatis mutandis*, to Elapsed Time. Isolating the attitude to Run Distance in this way differs from, but is not inconsistent with, the interpretation of Distance as a sort of Income effect, even if respective roles may hard to distinguish in practice.

The understanding of Distance as a sort of Income effect associated with Time arose in the context of DDF models of type (0) where researchers were having *extreme*, not to say *fatal*, difficulties obtaining expected signs on the coefficients of Time and Fare LOS variables used in each Modal demand equation.

To make the point with a simple linear case for two modes and Time, and remembering to define excess time between modes 1 and 2 as $EX = (Time_2 - Time_1)$, they were estimating equations with structure (A) in:

(18-F)
$$T_1 = \underbrace{\beta_{1,1}\, Time_1 + \beta_{1,2}\, Time_2}_{(A)} = \underbrace{\beta_{1,1}\, Time_1 + \beta_{1,2}\, (Time_1 + EX)}_{(B)} = \underbrace{(\beta_{1,1} + \beta_{1,2})\, Time_1 + \beta_{1,2}\, (EX)}_{(C)}$$

seemingly without realizing that (adjustments made for logarithmic specifications of (18-F)), the combination [Time$_1$, Time$_2$] in (A) is strictly equivalent to [Time$_1$, EX] in (C).

In such circumstances, the sign obtained for $\beta_{1,1}$ in (A) will depend on the sign of $(\beta_{1,1} + \beta_{1,2})$ in (C) and sometimes come out "unexpected" with 2 or more modes, even in very high quality samples: in intercity (Domencich & Kraft, 1970) and urban (Kraft, 1963) cases, unconstrained cross coefficients typically yielded *nonsensical* results for Times (and Fares) and all cross terms had to be constrained to secure at worst zero coefficients (which were duly obtained).

---

[91] In their meta-analysis, Mackie *et al.* (2003, Tables 12-13) notably studied interactions between Distance and Walk or Wait times with RP data. All power value estimates of Distance imply optimism but car users appear somewhat more optimistic than public transport (bus and rail) users.
[92] There is no use of Distance power terms in her previous complementary work (Ramjerdi & Rand, 1992).

In the aftermath of silly results obtained, format (0) consistently yielded "expected" signs and prospered over the years by retaining solely own-Time and own-Fare terms, imposing in effect diagonal slavery to keep out cross-terms: such is the IIA *coup d'état* magic practice for modes.

Table 11. Three laws of rail demand under certainty, risk, and distance attitude assumptions

---

A. Marshall's (1890) law of demand under Price certainty assumptions (UC)

"There is then one general *law of demand:*—The greater the amount to be sold, the smaller must be the price at which it is offered in order that it may find purchasers; or, in other words, the amount demanded increases with a fall in price, and diminishes with a rise in price. There will not be any uniform relation between the fall in price and the increase of demand. A fall of one-tenth in the price may increase the sales by a twentieth or by a quarter, or it may double them. But as the numbers in the left-hand column of the demand schedule increase, those in the right-hand column will always diminish." Book III, Ch. III.2.

can be spatialized and applied to rail Price and Speed, or to rail Fare and Time, with BCT. The Level-of-Service (LOS) slopes must:

(a) be negative, *i.e.* $\left\{\beta_{rail, X_{Fare}} < 0\right\}$, and by extension $\left\{\beta_{rail, X_{Time}} < 0\right\}$;

(b) exhibit "non uniformity" of response, *i.e.* curvature $\left\{\lambda_{rail, X_{Fare}} \neq 1\right\}$, and by extension. $\left\{\lambda_{rail, X_{Time}} \neq 1\right\}$;

and may be such that:

(c-1) their respective absolute values fall (in accordance with the second partial derivatives) either at a ***damped*** rate with $\left\{\lambda_{rail, X} < 1\right\}$, or at an ***amplified*** rate with $\left\{\lambda_{rail, X} > 1\right\}$;

(d-1) their relative values (the marginal rate of substitution or VOT) either fall in ***damped manner*** with $\left\{(\lambda_{rail, X_{Time}} - \lambda_{rail, X_{Fare}}) > 0\right\}$, or increase in ***amplified manner*** with $\left\{(\lambda_{rail, X_{Time}} - \lambda_{rail, X_{Fare}}) < 0\right\}$.

B. Marshall's conjectural distinction between attitudes towards risk (over time) and towards outcome

"But in estimating the present marginal utility of a distant source of pleasure a twofold allowance must be made; firstly, for its uncertainty (this is an *objective* property which all well-informed persons would estimate in the same way); and secondly, for the difference in the value to them of a distant as compared with a present pleasure (this is a *subjective* property which different people would estimate in different ways according to their individual characters, and their circumstances at the time)." Book III, Ch. V.7.

finds a general expression in Rank Dependent Utility (RDU) specifications where transformations of probability distributions are used under Risk assumptions (UR) with risk support devices. But such devices do not yet treat Elapsed Time or Run Distance as risk supports proper.

C. If, in a spatialized model, the attitude to outcomes is expressed by BCT applied to LOS variables, and the attitude to Distance (DA) is expressed by a simple power term parameter $\gamma_{Distance}$ **applied to Distance** interaction terms, the enriched slopes with respect to Time or Fare may be such that:

(c-2) their respective absolute values fall either at a ***damped*** rate with $\left\{(\gamma_{rail, Distance} + \lambda_{rail, X}) < 1\right\}$,

or at an ***amplified*** rate with $\left\{(\gamma_{rail, Distance} + \lambda_{rail, X}) > 1\right\}$;

(d-2) their relative values either fall ***in damped manner*** with

$\left\{(\gamma_{rail, Distance_{Time}} - \gamma_{rail, Distance_{Fare}}) + (\lambda_{rail, X_{Time}} - \lambda_{rail, X_{Fare}}) > 0\right\}$, or ***increase in amplified manner*** with

$\left\{(\gamma_{rail, Distance_{Time}} - \gamma_{rail, Distance_{Fare}}) + (\lambda_{rail, X_{Time}} - \lambda_{rail, X_{Fare}}) < 0\right\}$.

and **distinctions between gross and net attitudes towards Distance can be made**, due account taken of the interpretation of increasing $\gamma_{Distance}$ parameters values as indicators of reduced Distance optimism up to pivot point of neutrality ($\gamma_{Distance} = 1$), and of Distance pessimism beyond it.

But the unavoidable systematic reintroduction of cross terms to account for close substitutes in demand and the issue of Distance attitudes bring the problem to the fore again. These considerations argue for specifications of type (15-B), not (15-A), independently from their demonstrated superior collinearity performance already alluded to. In microeconomic analysis of consumer behaviour, the

demand for a good is a function of its unit Price (not of expenditures on it or on competing goods) and of Income level: the latter sustains basket size whereas the former determine basket mix.

A necessary clarification of (18-F) then consists in adding a reference "time-income" index, and preferably one that is not decomposable into the times of the modes, for instance Distance, combined with Speed (and Price) rates.

Towards more complexity in the specification of (0-B). If a *Distance attitude* term is further of interest, the restrictions on BCT found in the simple transliteration from Expenditure to Rate metrics in (14-B) are in principle lifted: the new problem is that of duly accounting for time and money constraints, say by a standard term *Y* for the latter and a term *D* for the former, and of adding the Distance attitude interactions.

It finally becomes possible to specify the vague expression (0-B) and imagine, in the absence of Risk attitude concerns, a combination of Rate specification and Distance attitude. If two modes are competing and we neglect the constant and other possible variables, all non separable representative Utility functions will become very complicated, for instance for mode 1 in a two-mode case:

$$(18\text{-}G) \quad U_1 = f[\beta_{1,1} D^{\gamma_{P1}} P_1^{(\lambda_{P_1})} + \beta_{1,2} D^{\gamma_{P2}} P_2^{(\lambda_{P_2})} + \beta_{1,3} D^{\gamma_{V1}} V_1^{(\lambda_{V_1})} + \beta_{1,4} D^{\gamma_{V2}} V_2^{(\lambda_{V_2})} + \beta_{1,5} D^{(\lambda_D)} + \beta_{1,6} Y^{(\lambda_Y)}].$$

One realizes from this formulation decomposing gross demand slopes with products of non linear functions applied to each LOS variable that many real life specifications are extremely restrictive even without LOS uncertainty and under IIA limitations. Those include impedances based only on Distance, as in many Gravity models well surveyed in Erlander & Stewart (1990), or linearly constructed generalized costs, as in Abraham's Probit and Logit "path choice Law" in France (Abraham & Coquand, 1961), and elsewhere. They implicitly embody strong assumptions as to attitudinal trade-offs and rates of substitution among outcomes.

Why worry about gross damping or amplification? But, coming back to simplicity, if demand falls with Distance, how much does it in fact matter that gross values imply amplification or damping, *i.e.* differing rates of falling, especially if little can be done by transport firms or infrastructure providers to modify the attitude to Distance? We will argue below that amplification or damping will influence the market share distance profile of HSR gain forecasts, with implications for passenger and revenue project forecasts, and that using gross values falling faster rather than slower matters indeed.

But we must turn our attention beforehand on how obtaining the very sizes and signs of gross and net slope curvature coefficients naturally requires dealing explicitly with the estimation of form as a model dimension of unavoidable import where, in particular, the blind conveniences of untested fixed forms, long suspicious in an environment of monotonic curvatures, can be expected to convince even less in the emerging environment of co-monotonic curvatures (17-B) or (18-D).

5.6. Attitude to response asymmetry over 50 years of transport Logit models

It may be asked in passing why it took so long for non linearity of explanatory variables of Logit models to arise in a scientific environment where the BCT is "the most used non linear transformation in econometrics" (Davidson & MacKinnon, 1993).

The Logistic curve in pre-transport garb, 1838-1960. The discoverer of the logistic curve, Belgian mathematician Verhulst (1838), used it to describe population growth (Verhulst, 1845), as did others independently later in the Unites States (*e.g.* Pearl and Reed, 1920, 1927) and elsewhere, including Canada, and for the same reasons. Their interest was its sigmoid shape defined on the $\exp(V_i)$ quantities, as it was for Berkson (1944), promoter of the binomial canonical form (BNL) in

the bioassay literature. In more recent times, even the most sophisticated derivations of the Logit model by some economists (*e.g.* Leonardi, 1982 or 1984) also tended to ignore asymmetry[93].

The inception of asymmetric logistic curves in transportation, 1961-1962. But in 1961, at the very beginning of applications to transport and of the recognition of "Abraham's Law" in French road assignment projects, Abraham effectively used as representative utility function the logarithm of a (linear) generalized cost variable because it was obvious that road path choice would be better represented in this way: if the constructed cost variable is reduced to a unique element, this RUF is the equivalent of setting the BCT at 0 in the Standard Box-Cox Logit. Also concerned with data and fit, Warner (1962) compared various LOS forms and retained the logarithmic one after due analysis of residuals in his urban example. It is unfortunate that these transport applications[94] are typically ignored in essays on the history of the Logit model often excessively focussed on the later waves of development of 1968-1970 and of 1977[95] (*e.g.* Cramer, 2003; Andersson & Ubøe, 2010).

The McFadden stream, from 1968 onwards. In contrast with the data-driven form work stream associated with Abraham and Warner, a subsequent work stream addressing strictly comparable problems, namely road *tracé* choice (McFadden, 1968 or 1976a) and mode choice (CRA, 1972; Domencich & McFadden, 1975), used from the beginning the linear LOS form as RUF workhorse[96]. This central tendency has remained stable to this day in a vast consulting industry in transport and beyond, with limited heroic and inefficient linear approximation adaptations consisting in having as many linear models as there might be LOS classes of Price, Time and Distance, to say nothing of socioeconomic variables. It would seem that the Davidson-MacKinnon claim applies more to Classic than to Logit regression where one hardly goes beyond exhausting and often incoherent linearity by segment sub-sample.

Barriers to BCT asymmetry determination. But it is unclear how long this attitude to asymmetry can hold in view of even more general approaches applied for instance in Model 40 of Table 10, and of the availability, among others, of freeware such as TRIO (Gaudry *et al.*, 2001) for aggregate and discrete data, available since 1993, and BIOGEME (Bierlaire, 2003, 2008) for discrete data.

As discussed above, the application of BCT to explanatory variables in Logit models does not raise the same difficulties of applications found in Classic models, where the formulation of the Likelihood function of a dependent variable that cannot be negative (that is censored) is not straightforward[97]. But new barriers, common to (6-A) and (6-B), reflect new difficulties of application that impose a computational cost: the guaranteed unimodality of the Likelihood function has been lost; statistical pitfalls lurk due to the re-parameterization effected by BCT ─ for instance the surprising necessity to compute conditional *t*-statistics[98] that are invariant to the scale of

---

[93] When the present author worked with Giorgio Leonardi during the summer of 1983, our concern was the behaviour of probability limits in Dogit and IPT cores defined in Section 7, because positive asymptotes of choice probabilities that differ from 0 or 1 primarily reveal modeller ignorance or lack of relevant data. In that context, we felt we could treat the shape of the response curve and the existence of thick tails between the limits as distinct dimensions. This viewpoint was later justified by demonstrations with urban data for Winnipeg (Laferrière & Gaudry, 1992) and with intercity data for Canada (Gaudry, 1990 or 1993), the last of which is documented in Table 18 as Model 46.

[94] The seminal paper by Abraham & Coquand (1961) should be used by historians of the Multinomial Logit (MNL) as its inception in transportation.

[95] Year of the derivation of the *log-sum* aggregator. For the long story, see Williams (1977).

[96] Except here and there for some *ad hoc* logarithmic transformations of a variable in the utility functions of transit modes, at the beginning often cars owned but more recently a LOS variable, as documented above and below.

[97] For a recent summary, see Gaudry & Quinet (2010).

[98] Likelihood ratio tests remain exact but are tedious to apply if there are no automatic step-wise like procedures implemented in the computer programmes. Among freeware programs that do not implement this routine, TRIO (Gaudry *et al*,, 2001) compute the required conditional *t*-statistics directly, in addition to elasticities and values of time; BIOGEME (Bierlaire, 2003, 2008) computes unconditional *t*-values but it is possible to perform an additional iteration conditionally upon the BCT estimates to obtain conditional *t*-statistics of the regression coefficients.

regressors instead of the usual unconditional values of other non linear models (Spitzer, 1984; Dagenais & Dufour, 1994) ─; and forgotten numerical precision issues on computers resurface when BCT become relatively large, positively or negatively.

Readers concerned with these plumbing issues might want to test for themselves the results obtained for the models of Table 7, 8 or 16 for which databases and fully documented freeware algorithms, listed in Table 12, are easily accessible. Note that sample elasticities (η_1) are provided above in Table 2 and Figure 2, as well as in Table 4, and weighted aggregate probability point elasticities (η_4) in Table 3.

Table 12. Downloadable databases and algorithms used for some models of Tables 7, 8 and 18

MODEL[99]	DATA BASE	REQUIRED ALGORITHM			RESULTS
		Total demand	**Modal split**		Listed or used in
		λ_U	λ_{Time} , λ_{Fare}		
from Tables 7and 18 - National intercity models (Domestic only, except 9)					
2 / 3	*VIA RAIL 1987		**PROBABILITY**		Table 3 / Table 8
0**	Canada 1976			**SHARE**	Figure 1
44, 46					Table 18
8		**LEVEL**			Table 7
9	Germany 1985				Table 7
from Tables 8 and 18 – Urban models					
22, 26 / 48	Santiago 1983-1985		**PROBABILITY**		Table 8 / Table 18

DOCUMENTATION OF EACH ALGORITHM	ESTIMATION PROCEDURES	PROGRAM USER GUIDE & DATABASES	SOURCE & COMPILED PROGRAMS (IBM PC)
LEVEL			
L-1 option: serial correlation	Tran *et al.*, 2008	Tran & Gaudry, 2008a	Tran & Gaudry, 2008b
L-2 option: directed correlation	Tran & Gaudry, 2008c	Tran & Gaudry, 2008d	Tran & Gaudry, 2008e
PROBABILITY			
P-2 to P-6 options for cores	Tran & Gaudry, 2009a	Tran & Gaudry, 2009b	Tran & Gaudry, 2009c
SHARE			
S-1 to S-5 options for cores	Tran & Gaudry, 2008f	Tran & Gaudry, 2009d	Tran & Gaudry, 2009e
QDF: *elasticities of Modal demand calculated according to equation (43-E) for [Total] ⊗ [Split] model pairs*			
20 pair options	Tran & Gaudry, 2008g	Tran & Gaudry, 2010a	Tran & Gaudry, 2010b

In QDF, elements of (43-E) are calculated using sample elasticities evaluated at means. But the algorithms allow more:				
Elasticities calculated	L-1	L-2	P-2/P-6	S-1/S-5
η_1 : Sample: $\left.\dfrac{\partial y}{\partial X}\middle/\dfrac{X}{y}\right|_{means}$	Sample	Sample	Sample	Sample
η_2 : Moment: $\left.\dfrac{\partial[m_o(y)]}{\partial X}\middle/\dfrac{X}{[m_o(y)]}\right|_{means}$	m_1 = expected value m_2 = standard error m_3 = skewness			
η_3 : Weighted aggregate			W. aggr.	
η_4 : Probability point ≡ η_3 (probability)			Pr. point	
η_5 : Percentage point ≡ η_1 (share)				Perc. point

  * This database is found in the QDF algorithm documentation Tran & Gaudry (2010a). All other databases are found in the documentation of the algorithms required to estimate the models. The Santiago 1983-1985 database is found in Tran & Gaudry (2009b) and the Canada 1976 and Germany 1985 databases are found in Tran & Gaudry (2008d and 2009b).

  ** Model 0 is not listed in a table but specified just after Figure 1 showing its estimated responses.

---

[99] Some of the models were estimated with earlier versions of the Level, Probability or Share algorithms implemented in Version 1 of TRIO available since 1993 and in Version 2 distributed as freeware since 2001 (Gaudry *et al.*, 2001).

6. Form knowledge benefits: signs of variables, model forecasts, other

6.1. The interpretation of signs under Beta-Lambda correlation

We observed above in the discussion of Table 3.B how use of optimal BCT forms could redress "incorrect" regression sign obtained in a linear or another specification. If BCT matter, it is then legitimate to extend that observation and to comment on their linkage to regression signs. We start with the casual empiricism often met in practice and argue that the interrelatedness of statistical correlation and form argues for a more formal approach that is not without epistemological dimensions. We then ask what difference BCT estimation makes, notably for HSR forecasts.

A. *Taking the beat out of the betas by maximization of expected signs?*

Analysts sometimes play a "trial and error" game with forms until they have found the "right" beta regression signs. This puts forms at the service of signs and takes the beat out of the betas by working one's way towards desired signs through maximization of the number of expected signs rather than of log-likelihood values. As such combinatorial exercises are performed for both components of demand framework (1)-(2), we discuss both Classical and Logit malpractices.

Classical regression. Under a classical formulation of type (6-A) adopted to explain the variability of airline fares, Borenstein & Rose (1994) list two variants, linear and logarithmic, partially reproduced in Table 13.A. As that table makes clear, only two (braided) variables, the least significant ones in statistical terms, keep their signs: the remaining three dummy variables expressing the nature of market competition, the core issue of the paper, change signs but are highly statistically significant under both alternate form conditions. The authors choose one (let the reader guess which!) solely on the basis of regression sign anticipations and without providing any common statistical measure between candidate results: who said empirical work[100] was no fun?

Table 13. Signs changing with the form of variables in transportation studies

A. Airline Price Dispersion, U.S.A. (RP data) Borenstein & Rose (1994); Model 2 from Table 3				B. Discrete Passenger Mode Choice, Paris (RP data) Gaudry (1985); Models from Table 3 (1976 data)		
Coefficients and *t*-statistics conditional on form				Elasticities and *t*-statistics conditional on form		
y variable :	Gini ticket price dispersion index			y variable : 6 modal choices	Car mode	
X variables...	Linear	Log-log	Optimal	X variables…	Linear	Optimal
[**...**] Monopoly	+0,154 (+4,81)	-2,169 (-5,27)	n.c. n.c.	[**...**] Car cost	-0,11 (-3,27)	-0,17 (-4,15)
Duopoly	+0,174 (+4,97)	-2,033 (-9,46)	n.c. n.c.	Parking cost, worker cat. 1,3	-0,01 (-3,23)	0,00 (2,76)
Large-duopoly	-0,022 (-2,77)	-0,117 (-0,21)	n.c. n.c.	Parking cost, worker cat. 2,4	-0,01 (-2,48)	0,00 (1,70)
Small-duopoly	-0,017 (-1,89)	-0,067 (-1,10)	n.c. n.c.	Cars per worker in household	0,28 (6,05)	0,33 (6,92)
Competitive	+0,172 (+7,16)	-1.807 (-6,98)	n.c. n.c.	Car time	-0,28 (-4,62)	-0,28 (-3,76)
Lambda (fixed) Log-likelihood	1,00 n.c.	0,00 n.c.	n.c. n.c.	Lambda (fixed and estimated) Log-likelihood	1,00 -911,82	0,50 -904,03
Comment: No log-likelihood values are reported by the authors who arbitrarily choose between models after stating that « the main qualitative results are robust to changes in functional form » (*sic*!).				**Comment**: RATP staff pointed out that much work had gone into obtaining the "right" negative sign for the two parking cost variables over a two-year period of model development.		

In this problem, formal BCT tests of the appropriate form might well have invalidated the arbitrarily preferred choice, perhaps by finding a BCT value between the linear and logarithmic cases, and

---

[100] The motto of preferred sign maximization is: «*Wenn es keine Spaß macht, dan ist es keine Arbeit*».

insignificant conditional *t*-statistics, or perhaps even by finding an optimal point not too far from their rejected choice...and *contrary to the authors' expectation* that competition increases price variability, supposedly "verified" here by the last variable in the arbitrary linear case.

Logit regression. With models of type (6-B), it is frequent to find in Modal utility functions some variables appearing non linearly alongside others appearing linearly, both forms adopted without the slightest stated statistical justification. For instance, the ANTONIN model for the Paris area (STIF, 2004) uses the logarithm of modal costs without providing reasons — perhaps sign problems or even the unstated desire to impose rough absolute Cost damping[101] (13-F) found in other cities?

Unfortunately, in the analysis of bargaining games between data and analysts (Leamer, 1978a, 1978b; Ley, 2006), too little attention has been paid to the difficulties of producing honest results with form endogeneity built into the formulation and the maximization of the Likelihood function.

B. The joint determination of statistical correlation and form

We wish to emphasise here the reasons why the BCT is more than a legitimate curve determination or fitting device and cannot be dissociated from the determination of statistical causality to the extent that, with non-orthogonal data (where non-orthogonality itself also depends on form, a matter of interest in experimental design[102]), form and statistical correlation are in fact jointly determined.

The importance of non zero simple correlations. But what is known in fact and practice should also be expected from theory because signs in multiple regression depend on both the variances (on their standard deviations)[103] of regressors and on their covariances (their unweighted simple correlations)[104], elements that change with the power used for variables. To see this, consider slopes b_2 and b_3 of the least squares regression plane relating observations on a vector Y to those on a constant X_1 and on variables X_2 and X_3, as found in textbooks (*e.g.* Johnston, 1984, p. 81), written simply in terms of the standard errors of the variables and of their simple pairwise ("linear" or "simple Pearsonian") correlation based on raw unranked data:

(19)
$$b_2 = \frac{r_{12} - r_{13} r_{23}}{1 - r_{23}^2} \frac{s_1}{s_2} \quad and \quad b_3 = \frac{r_{13} - r_{12} r_{23}}{1 - r_{23}^2} \frac{s_1}{s_3},$$

where, lower case letters denoting deviations from means calculated as $\left[x_k = (X_t - \bar{X}) \right]$, the sample *standard deviations* of Y, X_2 and X_3 are, respectively:

(20)
$$s_1 = \sqrt{\frac{\sum y^2}{T}}, \quad s_2 = \sqrt{\frac{\sum x_2^2}{T}}, \quad s_3 = \sqrt{\frac{\sum x_3^2}{T}};$$

and the simple pairwise correlations are defined as

(21)
$$r_{12} = \frac{\sum (yx_2)}{\sqrt{\left(\sum y^2\right)\left(\sum x_2^2\right)}}, \quad r_{13} = \frac{\sum (yx_3)}{\sqrt{\left(\sum y^2\right)\left(\sum x_3^2\right)}}, \quad r_{23} = \frac{\sum (x_2 x_3)}{\sqrt{\left(\sum x_2^2\right)\left(\sum x_3^2\right)}}.$$

If the regressors X_2 and X_3 in (19) are orthogonal, the correlation coefficient $r_{23} = 0$ and the multiple regression slope coefficients b_2 and b_3 coincide with the simple (positive or negative) pairwise

---

[101] This specification also imposes relative Cost damping (15-C), contrary to all findings of Table 8.

[102] To the extent that variables are jointly determined by analysts for an SP questionnaire, one would not expect the variables as shown in the questionnaire to be orthogonal but rather their values transformed by the most likely BCT powers to be orthogonal.

[103] The standard deviations in (20) are the positive square roots of the variances.

[104] The covariances equal the correlation coefficients in (21) divided by T and multiplied by their own denominators: it is therefore unweighted by the geometric mean of the second moments of the separate variables and weighted only by T.

correlation coefficients (21). But if they are not orthogonal, the presence of $r_{23} \neq 0$ can change the signs of b_2 or b_3: it becomes determined by a difference of terms, one of them a product of correlation coefficients, of perhaps opposite signs. All of these simple correlation coefficients are modified in hard-to-predict ways by BCT and by corrections to re-establish spherical residuals.

Correlations among triply transformed variables. To get a sense of these modifications, remember that, in the presence of Box-Cox transformations and of heteroskedasticity defined as in (6-D), all variables of (6-A), dependent and explanatory, may be rewritten as:

$$(22) \qquad X_{kt}^{***} = \left\{ \frac{X_{kt}^{(\lambda_{xk})}}{f(Z_t)^{1/2}} - \sum_\ell \rho_\ell \frac{X_{k,t-\ell}^{(\lambda_{xk})}}{f(Z_{t-\ell})^{1/2}} \right\}, \quad \rho_\ell \neq 0, \forall \ell \; ; \qquad \text{or as} \qquad X_{kt}^{**} = \left\{ \frac{X_{kt}^{(\lambda_{xk})}}{f(Z_t)^{1/2}} \right\},$$

$$\text{(A)} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(B)}$$

depending on whether serial autocorrelation, assumed to be produced by a stationary process of order ℓ (with $\rho_\ell \neq 0, \forall \ell$), is present or not. In the latter case (B), the variables are still doubly transformed. In the absence of heteroskedasticity, the number of transformation operations is reduced by one in each case, both illustrated shortly with demonstrated parameter sign changes:

$$(23) \qquad\qquad X_{kt}^{**} = \left\{ X_{kt}^{(\lambda_{xk})} - \sum_\ell \rho_\ell X_{k,t-\ell}^{(\lambda_{xk})} \right\}, \quad \rho_\ell \neq 0, \forall \ell \; ; \qquad\qquad \text{or in} \qquad\qquad X_{kt}^{*} = \left\{ X_{kt}^{(\lambda_{xk})} \right\}.$$

$$\text{(A)} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(B)}$$

Doubly transformed variables in Classic regression. The first example, illustrated in Figure 9, pertains to a monthly Transit demand equation for schoolchildren in Montreal[105] where variables are doubly transformed, as in (23-A).

Figure 9. Travel time, Income and autoregression parameter sign changes under different forms
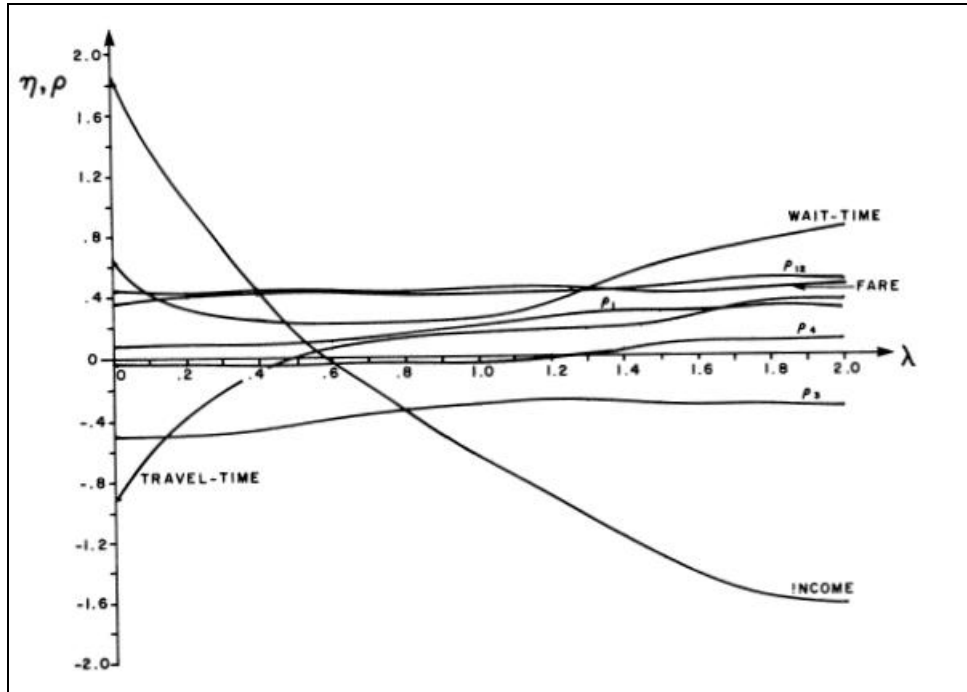

Figure 9 contains plots of some 4 demand elasticities (Wait time, Travel time, Fare, Income) and of 4 autoregressive parameters $(\rho_1, \rho_3, \rho_4, \rho_{12})$ drawn from this model and shows how elasticities of variables and autoregressive parameters vary with form, starting with the logarithmic case (at the

---

[105] Source: Gaudry & Wills, 1978, Figure 19.

origin) and ending with squared variables. We note sign changes for Travel time and Income around the square root point $\lambda = 0,55$ and a sign change for ρ_4 to the right of the linear case, at $\lambda = 1,30$.

Interestingly, one concludes with a logarithmic model that students travel more if speeds are lowered, and this despite the fact that transit is then an inferior good; but the reverse conclusion is reached if the model is assumed to be linear or the point of optimal form $\lambda = 0,84$ is accepted: longer travel times reduce demand and transit is a superior good. Should not the data decide?

Singly transformed variables in Logit regression. In the second example, reported in Table 13.B for the Car mode, the explanatory variables are transformed as in (23-B). In that respect, the demonstration resembles that of Table 3.B except that the reference work trip model, again pre-specified independently from the author of the form tests (Gaudry, 1985), is urban with 6 modes and obtains only one BCT.

The first column reproduces results selected from a linear model specified and estimated by RATP & Cambridge Systematics (1982) using very high quality network data and travel choice information from a carefully made 1976 survey of work trips in Paris (Moïsi *et al.*, 1981). Only a single BCT was applied to all strictly positive variables. As shown in the second column, the log-likelihood gains were dramatic (15 log likelihood points; one degree of freedom of difference) and the optimal BCT estimate stood exactly at mid-point between 0 and 1. That square root value modified some of the elasticities by about 20% but, more importantly, changed the signs of the parking cost variables for the two highest socio-economic categories of Paris area employees: perhaps are their chauffeurs all the more useful that parking prices are high?

Simple correlation made complicated. Coming back to reasons for sign changes, consider now the general expression for any correlation coefficient (*e.g.* r_{23}) of a regressor variable in (19) rewritten in accordance with the triple transformation of (22-A) in terms of starred variables:

$$(24) \qquad r_{23} = \frac{\sum (x_2^{***} x_3^{***})}{\sqrt{\left(\sum (x_2^{***})^2\right)\left(\sum (x_3^{***})^2\right)}} .$$

One could be forgiven for concluding from (24) that playing with forms to obtain desired slope signs is a game that is hard to be predictable about. Even in two-variable regression, to say nothing of K-variable regression, we do not expect signs to be manually predictable from (24) as a practical alternative to matrix inversion of transformed variables in (6-A), or its equivalent in (6-B).

But we do not exclude the possibility of further research illuminating the analytical structure of (24) with respect at least to the monotonic Box-Cox transformation which, in contrast with a simple power transformation X^γ, preserves the ordering of the data[106] and possesses two important invariance properties, namely

(25-A) $\qquad \beta_r \cdot \left[rX\right]^{(\lambda)} = \beta_s \cdot \left[sX\right]^{(\lambda)} = \beta \cdot \left[X\right]^{(\lambda)}$ *for any scalar r,s* \qquad [to a scalar transformation]

and

(25-B) $\qquad \left[X^\gamma\right]^{(\lambda)} = X^{(\gamma \cdot \lambda)} = X^{(\lambda)}$ *for any value of γ* \qquad [to a power transformation]

where (25-A) expresses invariance of the BCT to a scalar transformation of the data if there is a regression constant (Schlesselman, 1971) and (25-B) expresses invariance to a power

---

[106] Johnston (1984, p. 63) shows this vividly with two values of Y, namely 10 and $e = 2.84128$, as documented in Appendix A.

transformation of the data even in the absence of a regression constant (Gaudry & Laferrière, 1989) and involves a rescaling of the betas (neglected for simplicity).

We therefore argue that the interpretation of the size and sign of beta regression coefficients and of their derived statistics such as demand elasticities, requires explicit form tests. BCT form estimates can be easily reported and the trade-off between researcher priors and Likelihood values dependent on form transparently provided. Otherwise, results obtained from models of untested form remain suspiciously conditional.

C. Form sensitive causality: the Beta-Lambda correlation proposal

Beta-Lambda correlation or causality. In discussions of statistical causation within parametric models, well summarized in Fridstrøm (1992), the understanding of causality, even defined solely in terms of statistical association such as the basic approach introduced by Wiener (1956) and Granger (1969) to study dynamic relationships between and among time series, does not explicitly emphasize a role for functional form, be it the local approximation kind, such as Box-Cox, or a more global kind, such as Fourier. Statistical correlation is not explicitly form sensitive[107] and unconditional in that sense, despite the fact that some authors realize that they only "yield a complete picture of *linear* causality properties" (Dufour & Renault, 1998).

To address this need, and for want of a better expression in the circumstances, we call *Beta-Lambda correlation* (or causality, if one prefers) the joint data-based statistical establishment of regression coefficient and BCT power form[108] parameters, an issue formally addressed in Cho & Ishida (2012).

Factor & Form composition of systematic GLM components. Limiting the discussion to the confines of the Generalized Linear Model (GLM) framework introduced by Nelder and Wederburn (1972), it is clear that multiple dimension *Factor and Form (F & F)* construction of the systematic component poses special problems independent from the link and random component specification of the model.

For instance, the direct or inverse Box-Cox and Box-Tukey parameterization of variables in regression involves asymmetries in the sense that the value of Lambda does not matter independently from the value of Beta (*e.g.* it does not matter at all if Beta is zero) but the value of Beta matters even when the value of Lambda does not. We avoided above these limit statistical cases better left to specialized discussions and we concentrated on the credibility of some non linear form estimates in order to establish a first benefit of endogenous forms.

A first benefit of Beta-Lambda correlations. In our mind, it suffices that the establishment of correlations among variables that are not conditional on their *a priori* mathematical forms in multivariate regression, but are jointly determined with the forms themselves, add some flesh or structure to the "constant conjunction" which is said by David Hume to give rise to our sense of causality.

In particular, we do not see why measures of statistical causality should be restricted by assumed log-linearity or linearity of regressors in transport, or elsewhere: jointly unconditional estimates of

---

[107] A recent and representative example of the unconscious neglect of form is found in Dufour & Taamouti (2010) where the parametric theory part is written in terms of VARMA models with all variables appearing linearly and the empirical illustration with all variables suddenly appearing in logarithmic form. The door is implicitly closed on the possibility that a model could be cointegrated of order 1 say in logarithmic form but not in Box-Cox optimal form, as found by Djurisic (2000). Concerning cointegrated models, Granger (2010, p. 4) states: "Note, we often see X_t and log X_t being considered as modeling alternatives, with the assumption that they have the same time series properties." The same author just limits forecasting to time-series (Granger, 2008).

[108] For reasons explained in Appendix 1, it seems unwise to include simple power transformations in this definition.

form and regression parameters have at least this epistemological benefit and, as we shall presently see, other benefits as well because they make a practical and demonstrable difference to forecasts.

6.2. The comparison of forecasts made under different functional forms

After showing that non linearity makes a difference to the signs obtained for particular X_k variables, can one indeed demonstrate, for any variable X_k, that it also makes a numerical difference to model forecasts, defined as calculable values of dependent variables?

A. Two methods

To compare the forecasting ability of models, we consider the set of differences between values producible by a pair of selected model variants differing only in form[109] and:

(i) [*Resolution method*]: either characterize (without calculating all individual values) this set with respect to any variable X_k potentially present in both variants[110] (and perhaps suitably modified)[111], by finding some diagnostic points: the values of X_k that make the difference change sign (a crossing point A), reach a maximum (the point of maximum difference B) or induce a slope change (an inflexion point C);

(ii) [*Enumeration method*]: or characterize the same set by first calculating all individual values of the differences and then by studying the set of effectively calculated values against the X_k variable of interest (or possibly another), for instance by plots called *X-Profiles* of Δp_{X_1} at the beginning of Section 3.

If Enumeration is straightforward, finding those ABC points of diagnostic with respect to X_k by Resolution is sometimes, even for our cases differing only in form, another matter: it may be analytically impossible and require resort to numerical resolution, a quandary that differs between models of Total market size estimated by Classical regression and models of Mode choice estimated by Logit regression. In neither case does our presentation of the Resolution method assume that the reference variant is of particular form, but our applications of the Enumeration method with Logit models below all take as reference variant the popular linear case.

B. The Resolution method and forecasts by Classical regression models

At a common sense motherhood level, "when the necessary underlying assumptions are true, the Box-Cox transformation works well and does produce superior forecasts when a transformation is really justified" (Nelson & Granger, 1979), a "superior forecasting ability" that even holds in simultaneous equation models (Spitzer, 1977).

But can anything more systematic be said about the difference between forecasts from two Classical Box-Cox regression variants 1 and 2 of a model of type (6-A) differing here by assumption only in the constraints imposed on the λ applied to each? Consider such models and their difference:

(26) $$Y_{1n} = \beta_{10} + \sum_k \beta_{1k} X_{kn}^{(\lambda_{1k})} + u_{1n} \qquad \text{and} \qquad Y_{2n} = \beta_{20} + \sum_k \beta_{2k} X_{kn}^{(\lambda_{2k})} + u_{2n}$$

(27) $$\Delta Y_n = \left[\beta_{20} + \sum_k \beta_{2k} X_{kn}^{(\lambda_{2k})} \right] - \left[\beta_{10} + \sum_k \beta_{1k} X_{kn}^{(\lambda_{1k})} \right]$$

---

[109] It simplifies notation here that variants have the same explanatory variables but this condition is not necessary.
[110] Because the diagnostic can attribute differences with respect to any variable, any variable may be considered and it need not be present in both variants, but we focus here on variables that are.
[111] In the generation of forecasts, the levels of variables may be those from the sample but it is also possible to modify only one variable in both variants, as done with rail Time below, to understand its contributions to the difference in forecasts generated with that variable set at project scenario levels in both variants.

where indices $k = 1, ..., K$ apply to variables, $n = 1, ..., N$ refer to observations, β_{10}, β_{1k} and λ_{1k} are the parameters associated to variant 1, β_{20}, β_{2k} and λ_{2k} are those associated with variant 2 and the difference ΔY_n between values forecasted by the fixed parts neglects their error terms u_{1n} and u_{2n}. Table 14 presents the equations to be solved, and the analytical solutions when they exist, to locate points A, B and C characterizing the behaviour of difference (27) with respect to the variable X_{qn} of interest for the diagnostic.

Table 14. Analysis of differences between Total demand models differing in form values

(28-A)	To find a crossing point $\Delta Y_n = 0$, solve for X_{qn}^*:
	$$\Delta Y_n = \frac{\beta_{2q}}{\lambda_{2q}} X_{qn}^{\lambda_{2q}} - \frac{\beta_{1q}}{\lambda_{1q}} X_{qn}^{\lambda_{1q}} + \left[\beta_{20} - \frac{\beta_{2q}}{\lambda_{2q}} + \sum_{k \neq q} \beta_{2k} X_{kn}^{(\lambda_{2k})} \right] - \left[\beta_{10} - \frac{\beta_{1q}}{\lambda_{1q}} + \sum_{k \neq q} \beta_{1k} X_{kn}^{(\lambda_{1k})} \right] = 0$$
	where we note that the terms in brackets do not contain X_{qn}.
and	
(28-B)	To find the point of maximum difference $\partial \Delta Y_n / \partial X_{qn} = 0$, solve for X_{qn}^{**}:
	$$\frac{\partial \Delta Y_n}{\partial X_{qn}} = \beta_{2q} X_{qn}^{\lambda_{2q}-1} - \beta_{1q} X_{qn}^{\lambda_{1q}-1} = 0$$
	which yields $\qquad X_{qn}^{**} = \left(\frac{\beta_{1q}}{\beta_{2q}} \right)^{\frac{1}{\lambda_{2q}-\lambda_{1q}}}, \quad (\lambda_{1q} \neq \lambda_{2q} \text{ and } \beta_{1q}/\beta_{2q} > 0)$
and	
(28-C)	To find an inflexion point $\partial^2 \Delta Y_n / \partial X_{qn}^2 = 0$, solve the second derivative for $\tilde{X}_{qn}$:
	$$\frac{\partial^2 \Delta Y_n}{\partial X_{qn}^2} = \beta_{2q}(\lambda_{2q}-1) X_{qn}^{\lambda_{2q}-2} - \beta_{1q}(\lambda_{1q}-1) X_{qn}^{\lambda_{1q}-2}, \ (\lambda_{1q}, \lambda_{2q} \neq 1,2)$$
	which yields: $\qquad \tilde{X}_{qn} = \left[\frac{\beta_{1q}(\lambda_{1q}-1)}{\beta_{2q}(\lambda_{2q}-1)} \right]^{\frac{1}{\lambda_{2q}-\lambda_{1q}}}, \quad (\lambda_{1q} \neq \lambda_{2q} \text{ and } \frac{\beta_{1q}(\lambda_{1q}-1)}{\beta_{2q}(\lambda_{2q}-1)} > 0)$
	iff, as X_{qn} goes through the inflexion point, there is a change in sign of the second derivative.

Consider first (28-A) where the difference is found after isolating X_{qn} outside of the summation signs in (27) and developing the transformations $X_{qn}^{(\lambda_{1q})}$ and $X_{qn}^{(\lambda_{2q})}$. Note that, when $\lambda_{1q} \neq \lambda_{2q}$ as it must be for our purposes[112], a crossing point X_{qn}^* cannot be found analytically but only numerically, except for the very special quadratic cases [$\lambda_{1q} = 2$ and $\lambda_{2q} = 1$] or [$\lambda_{1q} = 1$ and $\lambda_{2q} = 2$], and need not be unique. But the point of maximum difference X_{qn}^{**} in (28-B) and an inflexion point $\tilde{X}_{qn}$ in (28-C) can be found analytically. One can also determine from the second derivative in (28-C) whether the point of maximum difference is a maximum or a minimum, depending on which of the following condition holds:

---

[112] We exclude $\lambda_{1q} = \lambda_{2q} = \lambda_q$ that is of academic relevance for us because our interest is precisely in $\lambda_{1q} \neq \lambda_{2q}$.

$$\text{(28-B Max)} \qquad X_{qn}^{**} < \left[\frac{\beta_{1q}(\lambda_{1q}-1)}{\beta_{2q}(\lambda_{2q}-1)}\right]^{\frac{1}{\lambda_{2q}-\lambda_{1q}}}, \quad (\lambda_{1q} \neq \lambda_{2q} \text{ and } \frac{\beta_{1q}(\lambda_{1q}-1)}{\beta_{2q}(\lambda_{2q}-1)} > 0) ;$$

$$\text{(28-B Min)} \qquad X_{qn}^{**} > \left[\frac{\beta_{1q}(\lambda_{1q}-1)}{\beta_{2q}(\lambda_{2q}-1)}\right]^{\frac{1}{\lambda_{2q}-\lambda_{1q}}}, \quad (\lambda_{1q} \neq \lambda_{2q} \text{ and } \frac{\beta_{1q}(\lambda_{1q}-1)}{\beta_{2q}(\lambda_{2q}-1)} > 0) .$$

It is therefore possible to demonstrate without excessive work by the Resolution method — finding ABC points — that different BCT form values assumed or estimated for X_{qn} must imply distinct numerical forecasts. One might of course want to perform an Enumeration analysis as well.

C. The Resolution method and forecasts by Logit regression models

As was already clear from Figures 1 and 2 on elasticities, one expects Logit forms to have impacts on revenue forecasts of HSR investments and, more generally, on reactions to significant changes in transport network conditions. But it would be important to have an alternative to the use of the Enumeration method and be more general to find out what lies behind different population elasticities (or values of time) that so conveniently summarize the behaviour of distinct models. To study market structure, the Ekbote-Laferrière revenue maximization exercise summarized above (see Table 4.C) relied only on Enumeration.

Problem formulation. Can anything systematic then be said with ABC diagnostic points, as we just did for models of the Total market, about the difference between forecasts from two Box-Cox Logit variants 1 and 2 of a model of type (6-B) differing by assumption only in the functional form structure applied to each?

Table 15 presents the equations to be solved and their solutions to locate points A, B and C for two such formulations with LOS transport variable X_{iqn} assumed present in the ith alternative of a multinomial ($i, m = 1,...,M$) Logit model:

$$\text{(29)} \qquad p_{1in} = \frac{\exp V_{1in}}{\sum_m \exp V_{1mn}} \qquad \text{and} \qquad p_{2in} = \frac{\exp V_{2in}}{\sum_m \exp V_{2mn}},$$

with n is an observation subscript and the ith representative utility components defined as:

$$\text{(30)} \qquad V_{1in} = \beta_{1io} + \sum_k \beta_{1ik} X_{1ikn}^{(\lambda_{1ik})} \qquad \text{and} \qquad V_{2in} = \beta_{2io} + \sum_k \beta_{2ik} X_{2ikn}^{(\lambda_{2ik})},$$

where the indices $k = 1,...,K$ again denote the independent variables and the $(\beta_{1i0}, \beta_{1ik}, \lambda_{1ik})$ and $(\beta_{2i0}, \beta_{2ik}, \lambda_{2ik})$ are the parameters associated to variants 1 and 2, respectively. The difference between forecasted shares of the ith alternative Δp_{in} due to changes in a variable $X_{1iqn} = X_{2iqn} = X_{iqn}$, the so-called own effect because we neglect here the cross-effect Δp_{jn} ($j \neq i$), is given by:

$$\text{(31)} \qquad \Delta p_{in} = p_{2in} - p_{1in} = \frac{\exp V_{2in}}{\sum_m \exp V_{2mn}} - \frac{\exp V_{1in}}{\sum_m \exp V_{1mn}}$$

Differences in forecasts: the general case $\left(\lambda_{1iq}, \lambda_{2iq} \neq 1\right)$. In Table 15, consider first crossing point to be determined by (32-A). For our purposes[113], when $\lambda_{1iq} \neq \lambda_{2iq}$, a crossing point X_{iqn}^* cannot

---

[113] We exclude $\lambda_{1iq} = \lambda_{2iq} = \lambda_{iq}$ that is academic for us because our interest is precisely in $\lambda_{1iq} \neq \lambda_{2iq}$.

be found analytically but only numerically, except for the very special quadratic cases [$\lambda_{1iq} = 2$ and $\lambda_{2iq} = 1$] or [$\lambda_{1iq} = 1$ and $\lambda_{2iq} = 2$]: the situation is therefore analogous to that of (28-A) above. However, unfortunately[114], in (32-B) and (32-C) cases, the solution cannot be found analytically, in contrast with (28-B) and (28-C) above. The critical point X_{iqn}^{**} that would solve (32-B) and the value $\tilde{X}_{iqn}$ that would solve (32-C) must both be found numerically. It is still possible to determine whether the point of maximum difference is a maximum or a minimum by considering the second derivative found in (32-C), evaluating it at the point X_{iqn}^{**}, and finding out numerically whether it changes signs when X_{iqn} passes through it.

Table 15. Analysis of differences between Logit Mode choice models differing in form values

(32-A)	To find a crossing point $\Delta p_{in} = 0$, we have to solve for X_{iqn}^*:
	$$\frac{\beta_{2iq}}{\lambda_{2iq}} X_{iqn}^{\lambda_{2iq}} - \frac{\beta_{1iq}}{\lambda_{1iq}} X_{iqn}^{\lambda_{1iq}} + R_{in} = 0 \ ,$$ and X_{iqn} is not contained in $R_{in} = -\log\left(\frac{S_{2n}}{S_{1n}Q_{in}}\right)$, $S_{1n} = \sum_{m\neq i} \exp V_{1mn}$, $S_{2n} = \sum_{m\neq i} \exp V_{2mn}$ or in $$Q_{in} = \exp\left\{\left[\beta_{2i0} - \frac{\beta_{2iq}}{\lambda_{2iq}} + \sum_{k\neq q}\beta_{2ik}X_{2ikn}^{(\lambda_{2ik})}\right] - \left[\beta_{1i0} - \frac{\beta_{1iq}}{\lambda_{1iq}} + \sum_{k\neq q}\beta_{1ik}X_{1ikn}^{(\lambda_{1ik})}\right]\right\}$$

and

(32-B)	To find the point of maximum difference $\partial \Delta p_{in}/\partial X_{iqn} = 0$, we have to solve for X_{iqn}^{**}:
	$$\frac{\partial \Delta p_{in}}{\partial X_{iqn}} = \left[\frac{\sum_{m\neq i}\exp V_{2mn}}{\exp V_{2in}^* \exp(\beta_{2iq}X_{iqn}^{\lambda_{2iq}}/\lambda_{2iq})} + \frac{\exp V_{2in}^* \exp(\beta_{2iq}X_{iqn}^{\lambda_{2iq}}/\lambda_{2iq})}{\sum_{m\neq i}\exp V_{2mn}} + 2\right]^{-1}\beta_{2iq}X_{iqn}^{\lambda_{2iq}-1}$$ $$- \left[\frac{\sum_{m\neq i}\exp V_{1mn}}{\exp V_{1in}^* \exp(\beta_{1iq}X_{iqn}^{\lambda_{1iq}}/\lambda_{1iq})} + \frac{\exp V_{1in}^* \exp(\beta_{1iq}X_{iqn}^{\lambda_{1iq}}/\lambda_{1iq})}{\sum_{m\neq i}\exp V_{1mn}} + 2\right]^{-1}\beta_{1iq}X_{iqn}^{\lambda_{1iq}-1} = 0, \ (\lambda_{1iq}, \lambda_{2iq} \neq 0)$$

and

(32-C)	To find an inflexion point $\partial^2 \Delta p_{in}/\partial X_{iqn}^2 = 0$, we have to solve for $\tilde{X}_{iqn}$ the second derivative:
	$$\frac{\partial^2 \Delta p_{in}}{\partial X_{iqn}^2} = \frac{\beta_{2iq}X_{iqn}^{\lambda_{2iq}-1}}{\left(A_{2n} + A_{2n}^{-1} + 2\right)}\left[(\lambda_{2iq}-1)X_{iqn}^{-1} - \frac{\beta_{2iq}X_{iqn}^{\lambda_{2iq}-1}\left(A_{2n}-A_{2n}^{-1}\right)}{\left(A_{2n}+A_{2n}^{-1}+2\right)}\right]$$ $$- \frac{\beta_{1iq}X_{iqn}^{\lambda_{1iq}-1}}{\left(A_{1n} + A_{1n}^{-1} + 2\right)}\left[(\lambda_{1iq}-1)X_{iqn}^{-1} - \frac{\beta_{1iq}X_{iqn}^{\lambda_{1iq}-1}\left(A_{1n}-A_{1n}^{-1}\right)}{\left(A_{1n}+A_{1n}^{-1}+2\right)}\right], \ (\lambda_{1iq}, \lambda_{2iq} \neq 0).$$ where $A_{1n} = \exp V_{1in}^* \exp(\beta_{1iq}X_{iqn}^{\lambda_{1iq}}/\lambda_{1iq})/\sum_{m\neq i}\exp V_{1mn}$, and $A_{2n} = \exp V_{2in}^* \exp(\beta_{2iq}X_{iqn}^{\lambda_{2iq}}/\lambda_{2iq})/\sum_{m\neq i}\exp V_{2mn}.$

The point of maximum difference is particularly interesting because determining its sign tells us something about the location of all other points in the sample, *i.e.* about whether one model always over-predicts relative to the other. But there is unfortunately no analytical solution for it, as expression (32-B) should make intuitively clear. In that expression, it has been possible to isolate variable X_{iqn} because, contrary to the previous writing of the expression for the derivative from

---

[114] As the interested reader may verify in Appendix 9 of Gaudry *et al.* (2007).

which it was derived, X_{iqn} does not appear in utility functions of other modes and is no more included in any of the $V_{1mn}, V_{2mn}(m \neq i), V^*_{1in}$ and V^*_{2in} functions. In that way, it is perhaps even more obvious that there is no analytical solution to determine the desired point of maximum difference X^{**}_{iqn}, as might also have been readily realized[115] by an analysis of that more complicated previous script just mentioned, namely:

$$(33) \quad \frac{\partial \Delta p_{in}}{\partial X_{iqn}} = \frac{\beta_{2iq} X_{iqn}^{\lambda_{2iq}-1}}{\dfrac{\sum_{m \neq i} \exp V_{2mn}}{\exp V_{2in}} + \dfrac{\exp V_{2in}}{\sum_{m \neq i} \exp V_{2mn}} + 2} - \frac{\beta_{1iq} X_{iqn}^{\lambda_{1iq}-1}}{\dfrac{\sum_{m \neq i} \exp V_{1mn}}{\exp V_{1in}} + \dfrac{\exp V_{1in}}{\sum_{m \neq i} \exp V_{1mn}} + 2} \quad .$$

The analytical proof of differences between forecasts generated from two model variants therefore requires harder work to find ABC points for Logit models than for models of Levels: it always requires numerical resolution. Alternatively, differences have to be studied by the Enumeration method after a proper sample enumeration (or calculation for all sample points) has been effected.

Differences in forecasts: the special case $\left(\lambda_{1iq} = 1, \lambda_{2iq} \neq 1 \right)$. But is this still true for point B if the reference case is linear as it often bound to be? In this typical case, (33) becomes:

$$(34) \quad \frac{\partial \Delta p_{in}}{\partial X_{iqn}} = \frac{\beta_{2iq} X_{iqn}^{\lambda_{2iq}-1}}{\dfrac{\sum_{m \neq i} \exp V_{2mn}}{\exp V_{2in}} + \dfrac{\exp V_{2in}}{\sum_{m \neq i} \exp V_{2mn}} + 2} - \frac{\beta_{1iq}}{\dfrac{\sum_{m \neq i} \exp V_{1mn}}{\exp V_{1in}} + \dfrac{\exp V_{1in}}{\sum_{m \neq i} \exp V_{1mn}} + 2} \quad .$$

where, given negative β_q slope coefficients and denominators that are necessarily positive, the sign of the maximum difference (34) still depends on the relative sizes of the two RHS terms. It remains obvious that the location of this potential turning point depends on more than the value of λ_{2q}, about which one can now ask a *further* question, before we proceed with Distance profile applications of the Enumeration method.

Special case profiles and damping or amplification. Is there a link between domains delineated by the pivot value $\lambda_{2q} = 1$ and the difficulty of finding diagnostic point B by the Resolution method?

It is clear from (34) that a damped or amplified value of λ_{2q}, despite its acknowledged role in establishing asymmetry of response, cannot be a sufficient condition to predict whether the point of maximum difference between two forecasts is negative or positive and, in either case, to determine as well if it is a true maximum rather than a mere unlikely fixed point.

By implication, neither could it be useful in predicting how many of the remaining n-1 points might stay on the other side of the x-axis from that of the location of the point of maximum difference. Indeed, why would the shape of Distance profiles, for example, be decisively determined by rates of change of the slopes (damping/amplification domains) rather that by the full model and the sample, including the size of X_{iqn}?

---

[115] This equation depends on two powers of the same variable X_{iqn}.

D. Using the Enumeration method on three Logit models: Distance profile analyses

After performing an enumeration analysis with respect to a certain variable, say Time, a Distance profile may be constructed by relating the results to Time itself (yielding a *Time-Profile* of Δp_{X_1}), or to another variable such as Distance itself (yielding a *Distance-Profile* of Δp_{X_1}). If each Time or Distance profile is unique, it is still possible to ask *three focussed questions* about any of them:

Q1. If a crossing point exists, is its existence independent from the double pivot $\lambda_T = 1$, as expected from our understanding of (34)?

This matters because $\lambda_T = 1$ is the pivot for slope independence (as distinct from amplification or damping) in (13-E) and for response curve symmetry (as opposed to asymmetry) in Figure 1.

Q2. As the existence of crossing points implies, with a linear reference model, a certain mix of linear model over-prediction (calculated $\Delta p_{in} > 0$ are below the x-axis) and under-prediction (calculated $\Delta p_{in} < 0$ points are above the x-axis), how much over-prediction occurs?

Q3. Is the amount of linear model over-prediction sufficient to imply excessive linear revenue forecasts? *i.e.* is it sufficient for the **maintained** non linear model to imply lower total trip and revenue forecasts than the **reference** linear model?

For a given over-prediction profile, the answer depends on whether, over the complete sample, linear model over-prediction is concentrated in relatively high total market size ranges or not. As total market size falls rapidly with Distance, the amount of over-prediction of relatively short trips is critically important to the existence of this linear model excess revenue bias.

We answer those questions for three maintained model examples of optimal non linear form and, in the Quebec-Windsor Corridor example, also for maintained models of non-optimal *a priori* forms.

i) Maintained optimal forms: Germany, the Pyrenees and the Quebec-Windsor Corridor

For the first three models, the optimal Box-Cox form structures involve the transformation of more than one variable, but we only graph results for a particular λ_{2q}, respectively $\lambda_{2,\text{Price}}$ for freight flows and $\lambda_{2,\text{Time}}$ for passenger flows, in Figures 10, 11 and 12.

ICE trains in Germany. Figure 10 from Mandel *et al*. (1994, 1997), was produced by comparing Inter City Express (ICE) train market share gains across an ICE scenario for Germany with a model of the whole country (Mandel *et al*., 1991) notably using for the Mode choice an extensive database from which 6 000 observations were drawn for estimation purposes.

After individual probability differences were calculated with linear and optimal non linear forms of the same intercity model for three modes, a best fit curve was estimated to summarize the structure of the cloud of individual Δp_{in} points, an S-shaped form where differences depend on OD distance or trip length. Within the sample range, a crossing point occurs around 150 km, an inflexion point around 350 km, and a maximum difference point at about 650 km.

The mix of linear model over-prediction for short trips and under-prediction beyond 150 km is clear but whether it implies linear model excess revenue bias[116] cannot be visually ascertained with certainty because the proportion of long trips required to offset short trips is not provided.

---

[116] The authors of Model 40 in Table 8 disregarded their superior Box-Cox tests and based passenger forecasts for an automatic train (called Orly-Val) accessing Orly airport from a suburban line on their linear results. Actual ridership for Orly-Val proved to be very inferior to linearly forecasted values and the case gave a bad reputation to Logit models.

Figure 10. Stylized Distance profile for ICE train scenario, Germany; 6000 obs. ($\lambda_T = 0,24$)
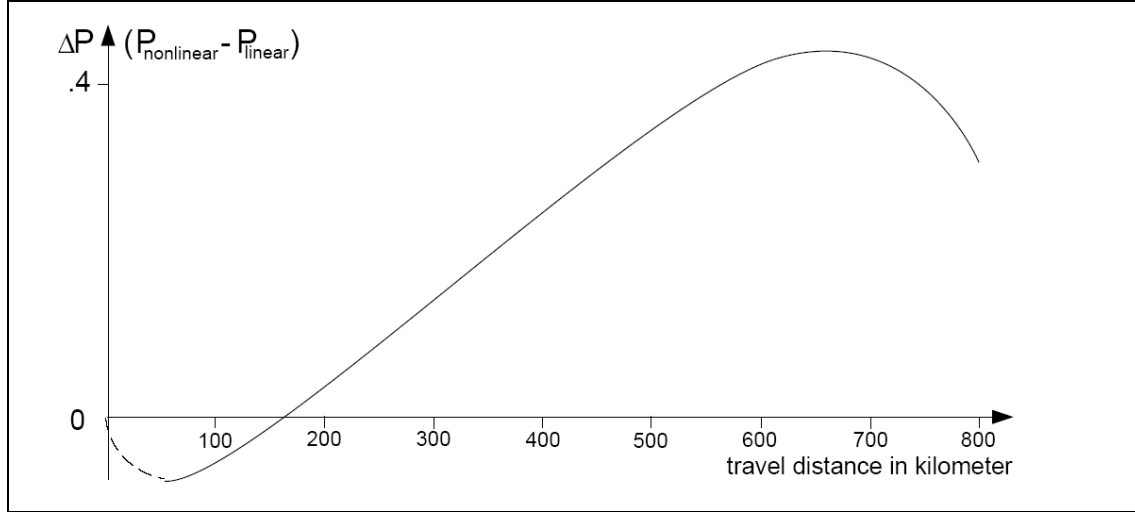
Intermodal trains across the Pyrenees. The only other extant example of the behaviour of Δp_{in} comes from the aggregate (share) freight model of continental European flows referred to above (Gaudry *et al.*, 2008), from which Equations (26) to (33) are also taken. In Figure 11, we use the Standard Box-Cox form results[117] graphed against road distance.

The differences shown in Figure 11 are calculated after decreasing intermodal train prices everywhere by 10% to determine the effect on all trans-Pyrenean intermodal train flows, within a model where classical trains and trucks are competing for traffic with intermodal container services.

The S-shaped structure resembles that of Figure 10, at least in the sense of some lower market shares at short distances and of large linear under-prediction at mid-range distances.

Figure 11. Distance profile for Intermodal train scenario, Pyrenean border; 749 obs. ($\lambda_P = -1,83$)

---
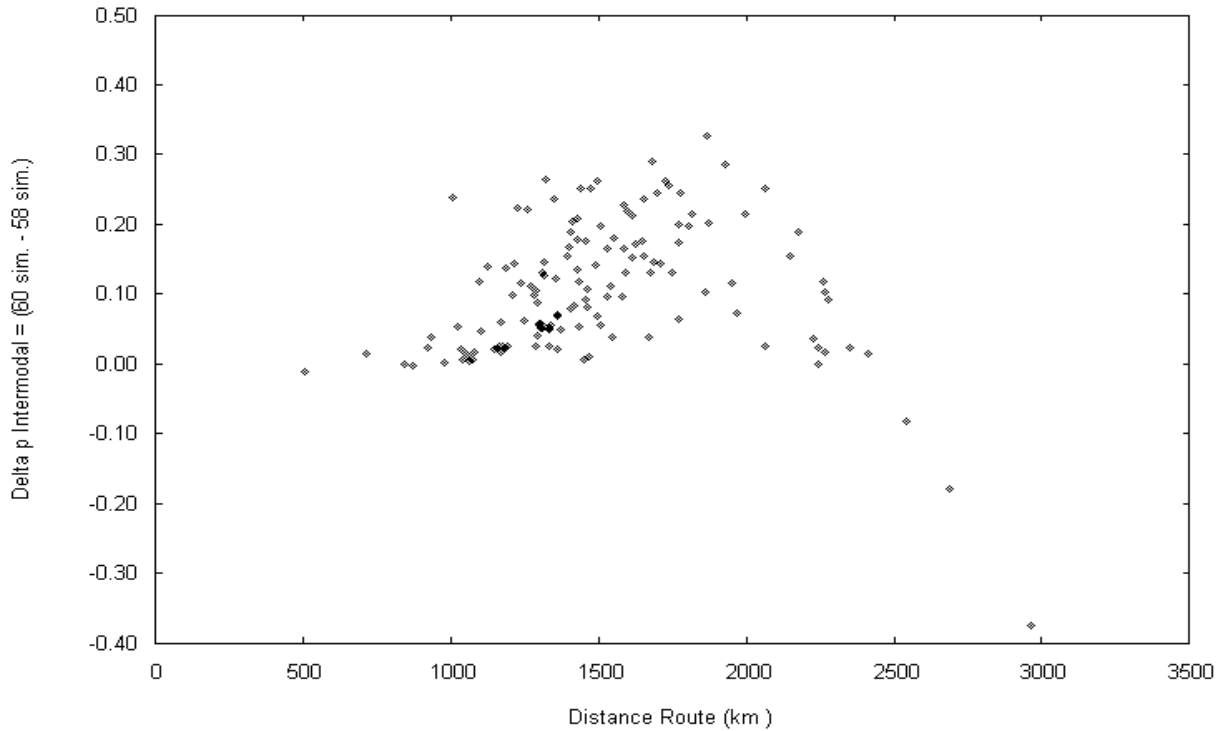
[117] The Generalized Box-Cox form, where the three modal prices appear in all Modal utility functions, is used by the French ministry for Transport since 2006.

HSR trains in the Quebec-Windsor Corridor. For Figure 12, we use the Business market models documented in Table 3.A and focus, as was done for Germany, on improvements in rail Time, comparing again the differences Δp_{in} between various non linear form predictions to a reference prediction made with the Linear model of Column 1. We will successively use as maintained model the optimal model and a series of non optimal variants.

For this trip purpose, rail is available for 4291 of the 4402 individuals in the sample and we assume that HSR implementation would reduce rail door-to-door Time by 50% for all. As exact OD distances are not provided in the final *official* database, we plot results against pre-project door-to-door rail Travel time (called Train own time in Table 3). Rail trip duration, which varies between 0,75 hours (45 minutes) and 6,75 hours (405 minutes) in the sample, is in any case very (positively) correlated with Distance.

Figure 12. Duration profile, HSR train scenario, Quebec-Windsor Corridor; 4291 obs. ($\lambda_T = 1,80$)
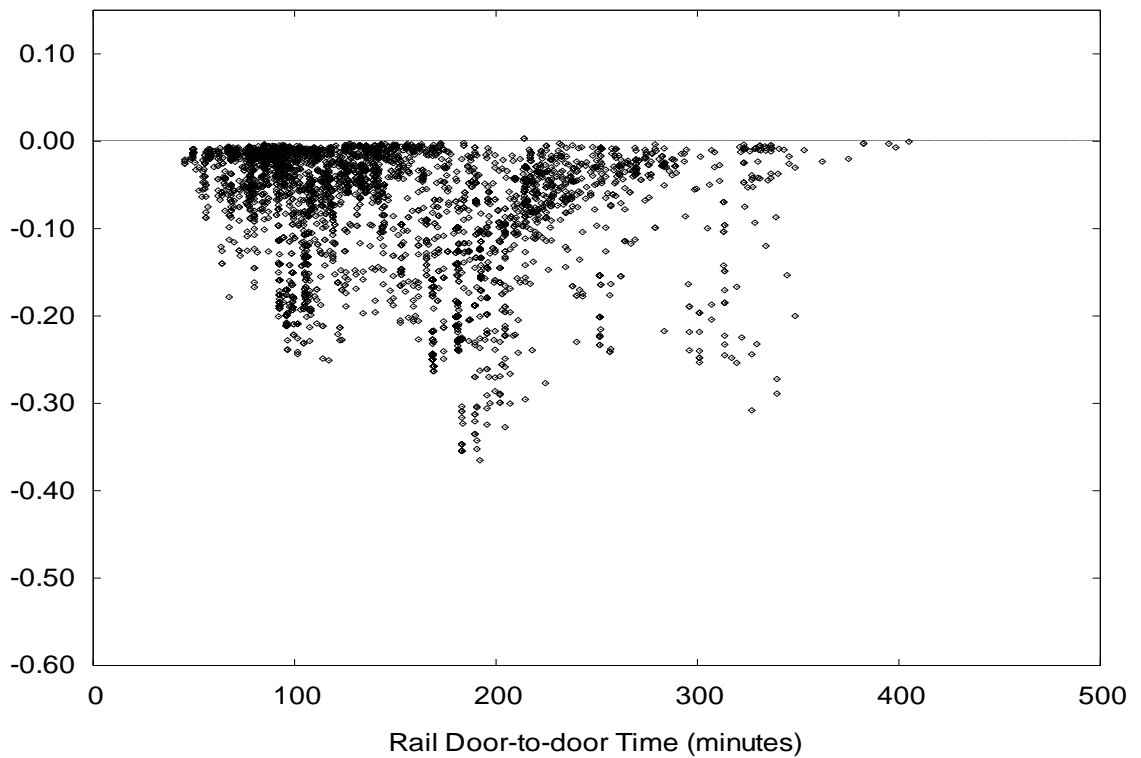
Despite the appearance of linear over-prediction everywhere, an exact count of the proportion of positive values of Δp_{in} finds traces of under-prediction (0,05%) at 3.5 hours (210 minutes), as indicated in the greyed column of Table 16, but this very small percentage shows as 0 in Figure 13.

Three answers for optimal form models. The questions raised can be answered in turn as follows:

A.1.1. *Crossing points and the pivot point*. The first two of the 3 maintained estimates imply damping and the last one amplification, but all $\lambda_{2q} = (0,24 ; -1,83; 1,80)$ produce some crossing points: the presence of crossing points is in practice independent from the side of the pivot one happens to be on. Is this the case?

One is misled by Figures 1, 10 and 11 in thinking that linear model under-prediction only happens with $\lambda_X < 1$: in Figure 12, linear model under-prediction also occurs where $\lambda_X > 1$.

A.2.1. *How much linear model over-prediction?* Linear model over-prediction is rare and too small to appear in Figure 12, but pervasive in Figures 10 and 11 at relatively short distances.

A.3.1. Enough linear model over-prediction to imply *excessive linear revenue forecasts*? Linear revenue over-prediction is almost everywhere in Figure 12 but concentrated on short trips in Figures 10 and 11. In the case of Germany, actual calculations showed that excessive linear revenue forecasts prevailed. In Figure 11, the situation is ambiguous but suggests the opposite.

**ii) <u>A series of maintained non optimal forms for the Quebec-Windsor Corridor</u>**

What happens if alternate assumptions about the BCT value of Time are used? It is important to realize that, in this new series studying for the Quebec-Windsor corridor the impact of alternate non optimal assumptions concerning the value of λ_{2q} on the Distance profile, λ_{2iq} does not vary in a *ceteris paribus* manner from the optimal[118] value, holding the rest of the estimated (form structure and regressor) parameters constant. Rather, for each postulated value of λ_{2iq}, all other parameters are re-estimated under the chosen constraint $\lambda_{2,\text{Time}}$ and each variant obtains its own log-likelihood value, as reported in Table 16 which deserves preliminary comments.

Table 16. Values of $\lambda_{2,\text{Time}}$, crossing points and $(\Delta p_n > 0)$ frequency in the Corridor (business)

Column	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
	Amplification $\lambda_T > 1$		Pivot	Damping $\lambda_T < 1$				
λ_T	1,80	1,20	1,00	0,00	-0,75	-0,80	-0,85	-1.80
Δp_{in}								
Crossing points $\Delta p_{in} = 0$	3,5 h	None	1,5 h	None	4,0 h	2,0 h	1,5 h	1,5 h
Range[119] of $(\Delta p_n > 0)$	*Small*	None	1,5-3,0 h	None	4,0-6,75 h	2,0-6,75 h	1,5-6,75 h	1,5-6,75 h
$\% (\Delta p_n > 0)$	*0.05%*	0%	1.82%	0%	2.00%	2.73%	3.75%	4.10%
$\mu (\Delta p_n)$	-0,08	-0,27	-0,14	-0,27	-0,26	-0,26	-0,26	-0,26
Log-likelihood	-1058.382	-1059.030	-1063.815	-1071.52	-1078.086	-1078.055	-1077.990	-1075.700

Statistical comment: multiple maxima and crossing points. By definition, the optimal Model listed in Column 1 of Table 16, shows the greatest log-likelihood gain as compared to the linear reference (from -1068,851 to -1058,382 in the comparison of Columns 1 and 2 of Table 3): it is therefore the "most likely" model in a statistical sense. We are now concerned with the values that would have been obtained for the Distance profile if the BCT of Time had been fixed *a priori* but the BCT on other variables (applied to Cost, Frequency and Income) had still been optimally estimated. As indicated in Table 16, the resulting log-likelihood function is quite flat with respect to the Time dimension BCT and even has a local maximum at Column 5. More information is provided in Figure 13.

Figure 13. The changing proportion of positive Δp_{in} (implying linear model under-prediction)

---
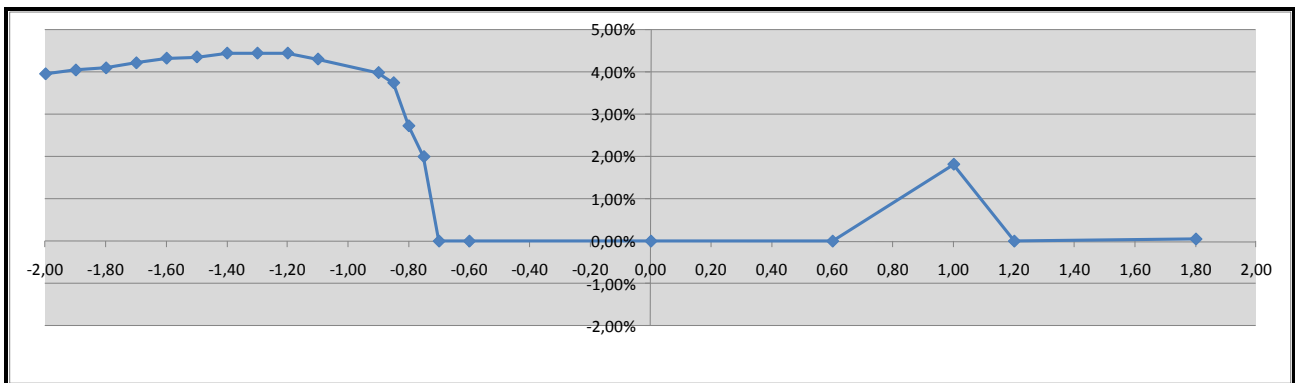
[118] We are not exploring the behaviour of Δp_{in} in the immediate vicinity of the optimal case, but across new variants.

[119] In the sample, the minimum rail travel time is 0,75 h (45 minutes) and the maximum is 6,75 h (405 minutes). This 0,45-6,75 range is unchanged in all columns because all cases are estimated with the full sample of 4291 observations.

Numerical comment: the frequency of linear over-prediction $(\Delta p_n < 0)$**.** In Figure 13, the percentage of positive Δp_{in} varies as one modifies *a priori* the maintained Time BCT value. Reading from right to left, the percentage rises at local maximum points 1,80 and 1,00 and monotonically increases beyond -0,75 despite an unchanged mean ($\overline{\Delta p_{in}} = -0,26$ in Table 16).

Three answers for non optimal form models. The questions raised can be answered as follows:

A.1.2. *Crossing points and the pivot point.* Are asymmetry and amplification systematically related to the existence of linear model over-prediction $(\Delta p_n < 0)$? One can see in Table 16 that over-prediction can occur on either side of $\lambda_T = 1$. It happens there over part of the range of amplification values ($\lambda_{2iq} > 1$ does not always over-predict) and over part of the range of damped values ($1,00 < \lambda_{2iq} < -0,75$). And neither does linearity imply the absence of positive Δp_{in}. Overall, the presence of $(\Delta p_n > 0)$ is independent from the double pivot point $\lambda_T = 1$, as it was for optimal form models.

A.2.2. *How much linear model over-prediction*? We now examine, between the optimal point $\lambda_{2iq} = 1,80$ and its opposite, $\lambda_{2iq} = -1,80$, three domains of assumed values successively yielding a small bulge of positive Δp_{in} at $\lambda_{2iq} = 1,00$, a monotonic emergence of positive Δp_{in} at $\lambda_{2iq} = -0,75$ and a leftward shift of the crossing point between $\lambda_{2iq} = -0,75$ and $\lambda_{2iq} = -0,85$. Linear model over-prediction is therefore sensitive to the value hypothesized for λ_{2q}.

The small bulge of positive Δp_{in} at $\lambda_{2iq} = 1,00$ for trips lasting between 1,5 and 3,0 hours. As one moves away from the optimal case presented in Figure 12, the distribution of values of Δp_{in} is profoundly modified, a new pattern without any positive values establishing itself both left and right (Figure 14 at $\lambda_{2iq} = 1,20$ resembling the logarithmic case and Figure 16 at $\lambda_{2iq} = 0,00$), with identical means at -0,27 of the pattern for the Linear case (Figure 15). The latter has both a slightly higher mean at -0,26 and some positive observations in relatively small numbers, hence the "small bulge" label. This bulge occurs for trip durations between 1,5 hour (90 minutes) and 3,0 hours (180 minutes) and the S-shapes profile resembles that for Germany (Figure 10).

The monotonic emergence of positive Δp_{in} at $\lambda_{2iq} = -0,75$ for trips of 4 hours or more. As one moves away from the representative logarithmic case (Figure 16), there are no positive values of Δp_{in} generated until one reaches $\lambda_{2iq} = -0,75$: market share gains then emerge for relatively long trips (Figure 17) lasting 4 hours (240 minutes) or more.

The leftward shift of emergence between $\lambda_{2iq} = -0,75$ and $\lambda_{2iq} = -0,85$ towards 1,5 hours. As one moves away from $\lambda_{2iq} = -0,75$, the emergence continues with a leftward shift in the minimum duration from trips lasting 3 hours (180 minutes) or more at $\lambda_{2iq} = -0,80$ (Figure 18) to trips lasting 1,5 hours (90 minutes) or more at $\lambda_{2iq} = -0,85$ (Figure 19). Moving towards more damped values of λ_{2q} has no effect on the crossing point of 1,5 hours or even on the visual appearance of plots (which are not shown) even as the proportion of positive differences Δp_{in} continues to increase to 4,5% around -1,45 and then falls slowly beyond.

A.3.2. Enough linear model over-prediction to imply *excessive linear revenue forecasts*? Most certainly, all profiles involve linear model over-prediction because significantly higher Δp_{in}

predominantly tend to occur for relatively long trips. In any case, the proportion of positive differences remains small: 4,10% at $\lambda_{2iq}=-1,80$ and 3.96% at squared value $\lambda_{2iq}=-2,00$.

Again, the pivot $\lambda_{2iq}=1,00$ does not matter in practice, but assumed BCT values do: results are consistent with the optimized Eckbote-Laferrière finding based on $\lambda_T=0,562$ (Table 4.A, Col. 2).

Figure 14. Corridor Time profile for HSR train scenario, constrained form ($\lambda_T=1,20$)
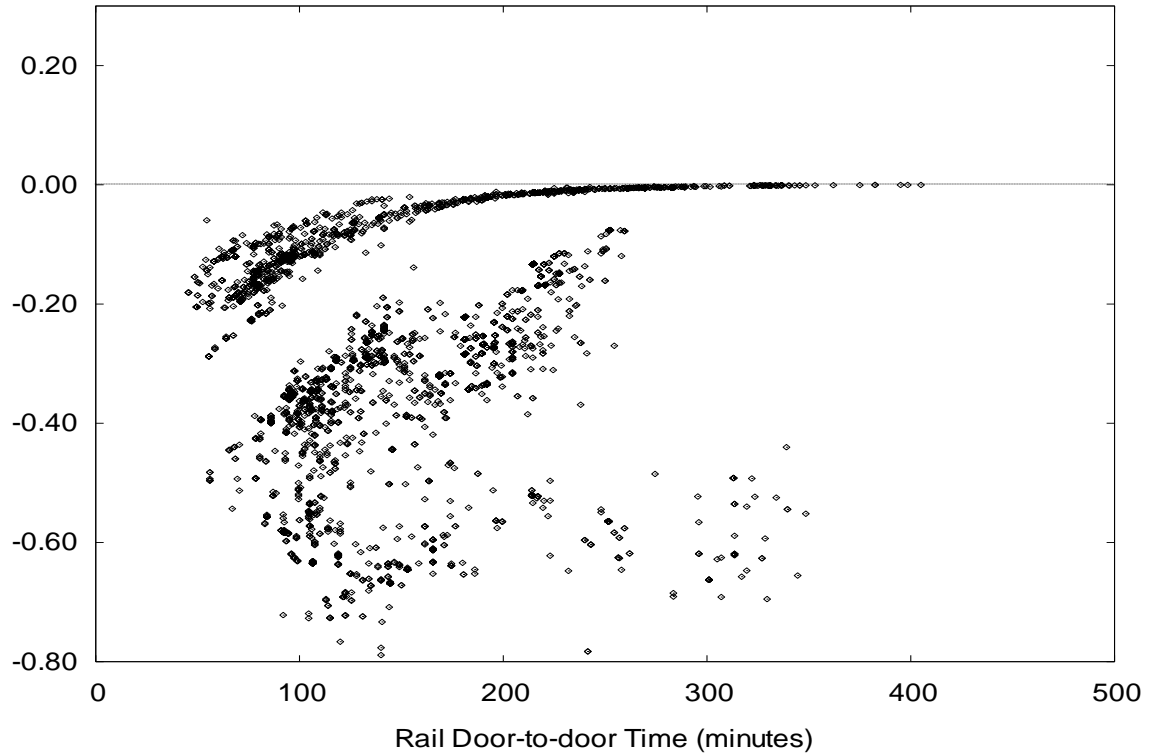
Figure 15. Corridor Time profile for HSR train scenario, Linear-constrained form ($\lambda_T=1,00$)
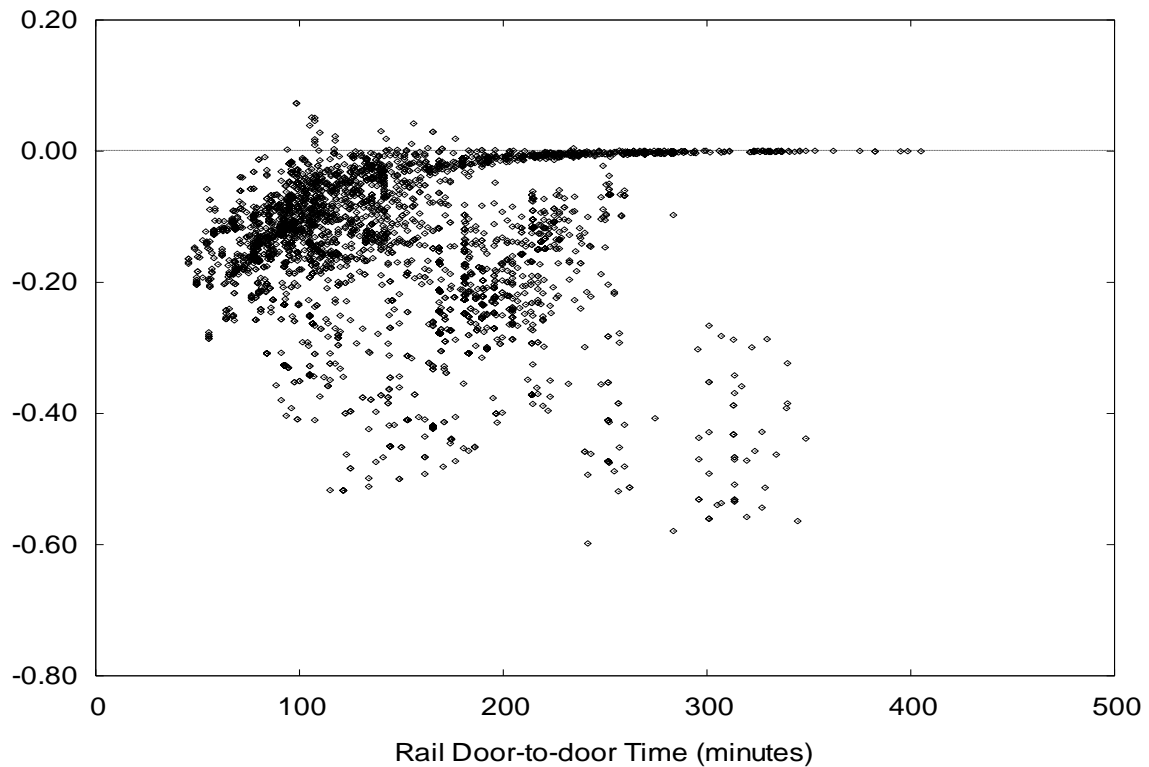
Figure 16. Corridor Time profile for HSR train scenario, Log-constrained form ($\lambda_T = 0,00$)

X-axis: Rail Door-to-door Time (minutes)
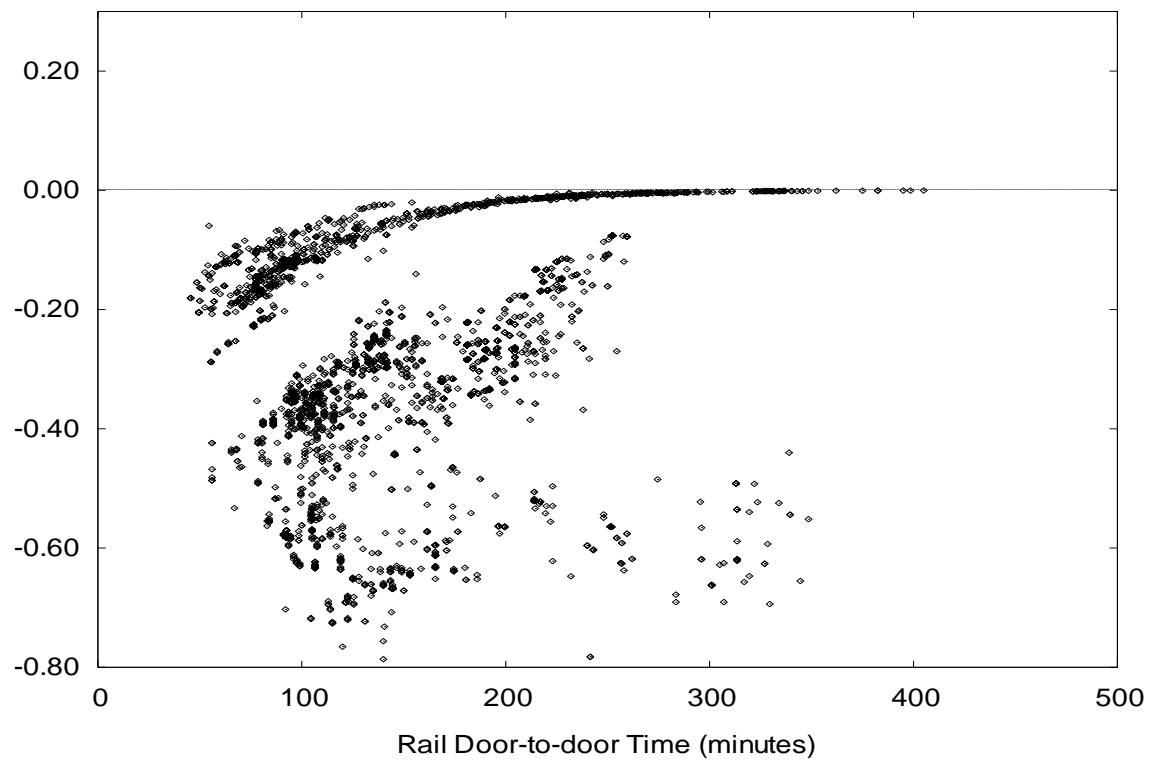
Figure 17. Corridor Time profile for HSR train scenario, constrained form ($\lambda_T = -0,75$)

X-axis: Rail Door-to-door Time (minutes)
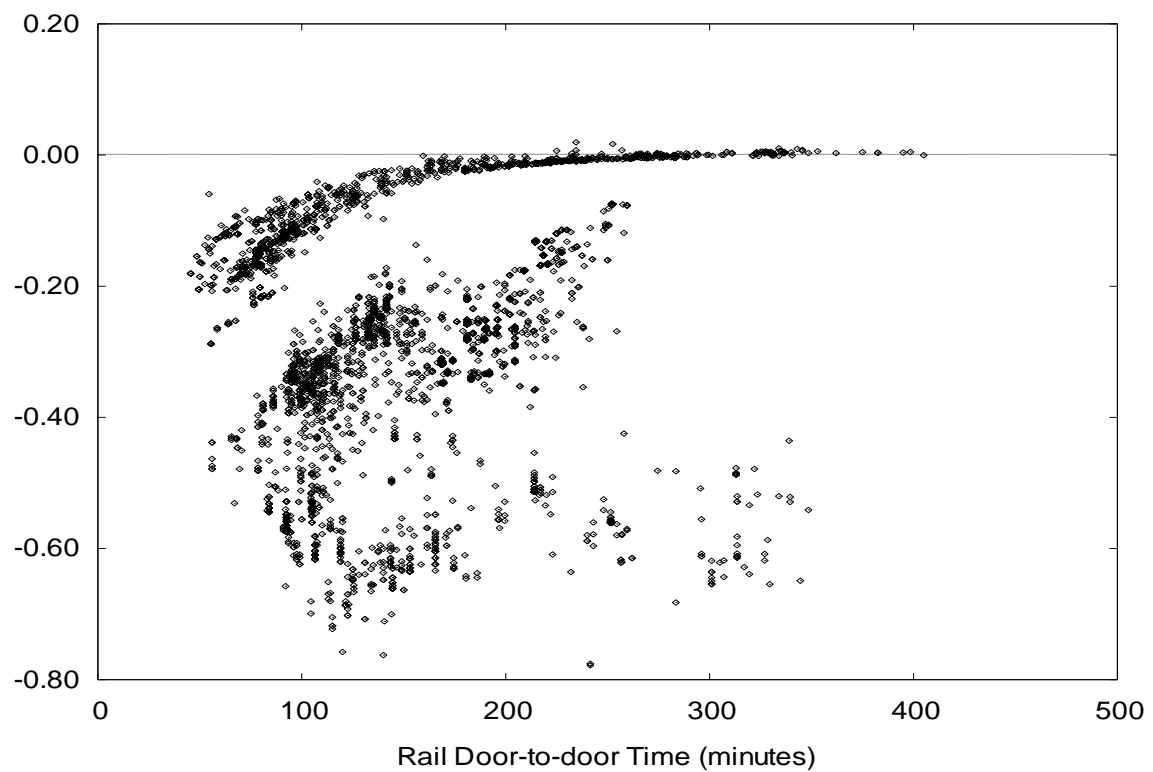
Figure 18. Corridor Time profile for HSR train scenario, constrained form ($\lambda_T = -0,80$)

Rail Door-to-door Time (minutes)
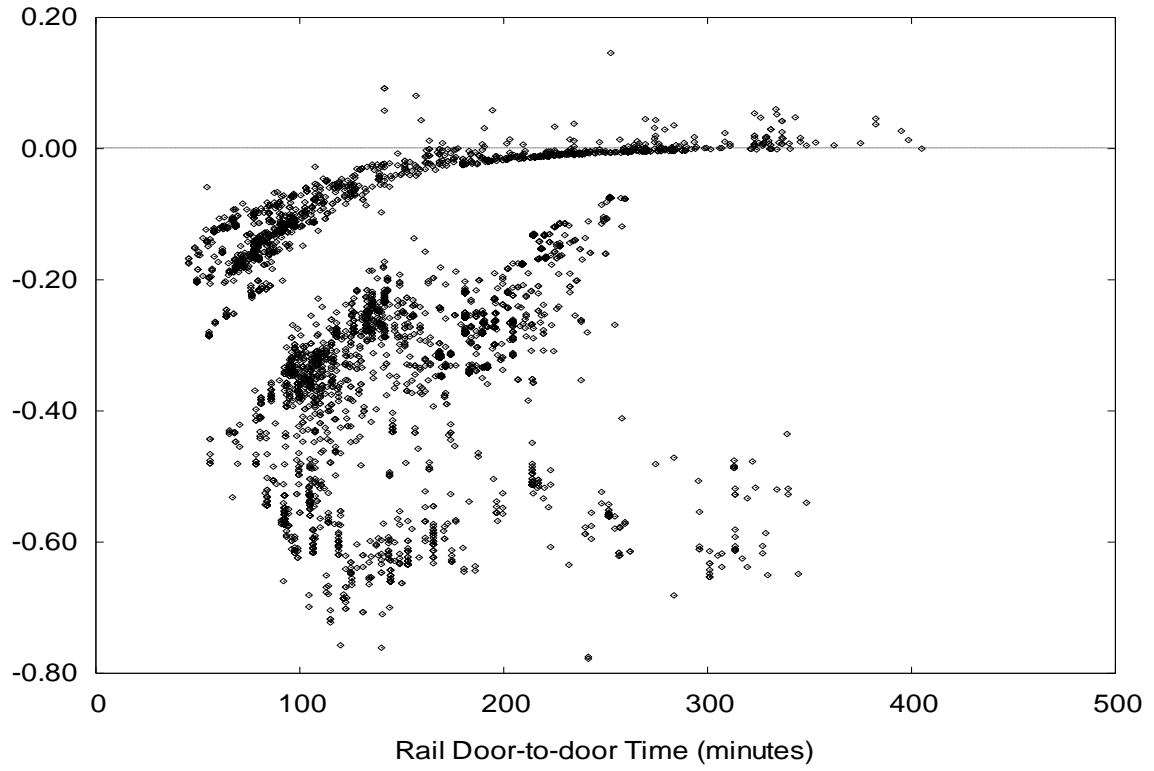
Figure 19. Corridor Time profile for HSR train scenario, constrained form ($\lambda_T = -0,85$)

Rail Door-to-door Time (minutes)
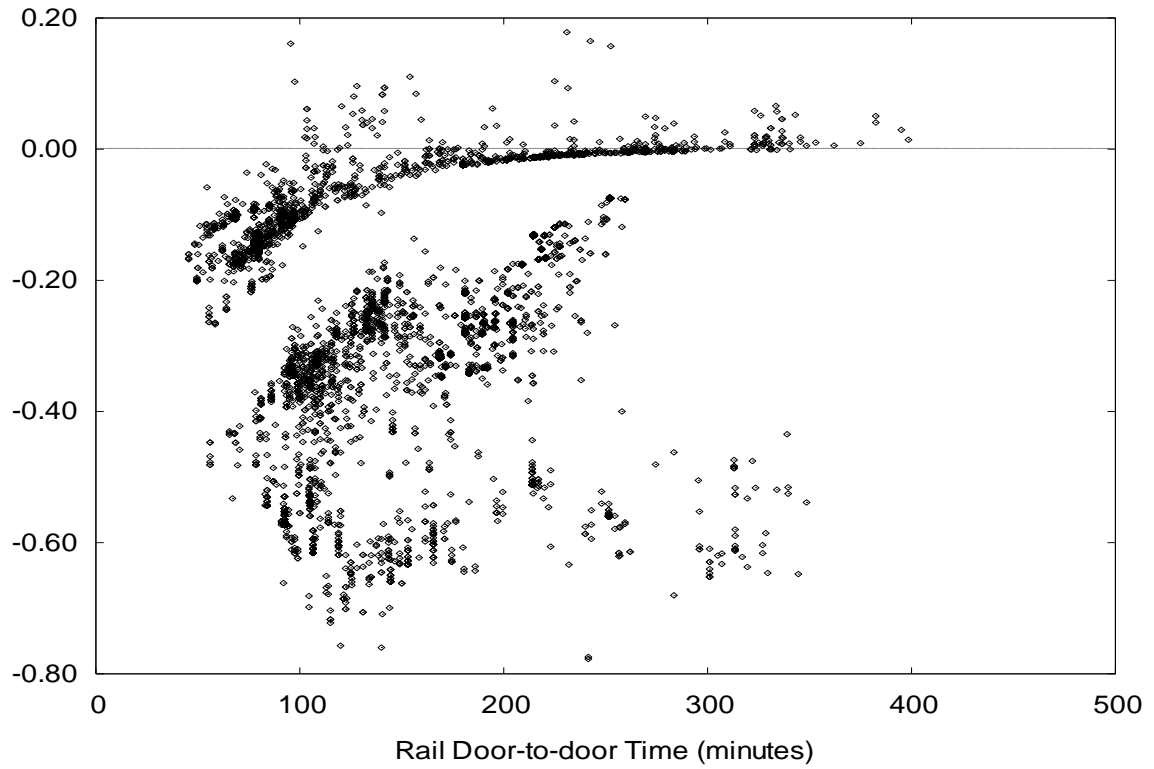
Quality of the VIA RAIL 1987 database. The results just shown are obtained with a high quality public database and do not depend on complex sampling schemes, as one can get an impression of from Figure 20 and Figure 21. In the former, one finds sampled business trips plotted within the 136 x 136 Origin-Destination matrix of zones defined for the Corridor. In Figure 21, the population values reconstructed from the representative weights of individual observations give the non-zero Rail Origin-Destination matrix flows (1719 out of 18496 OD pairs). It would be difficult to do tests with a better designed database (further details are given in Appendix C).

Figure 20. Spatial distribution, 136 x 136 Corridor OD matrix, 4401 business trip observations
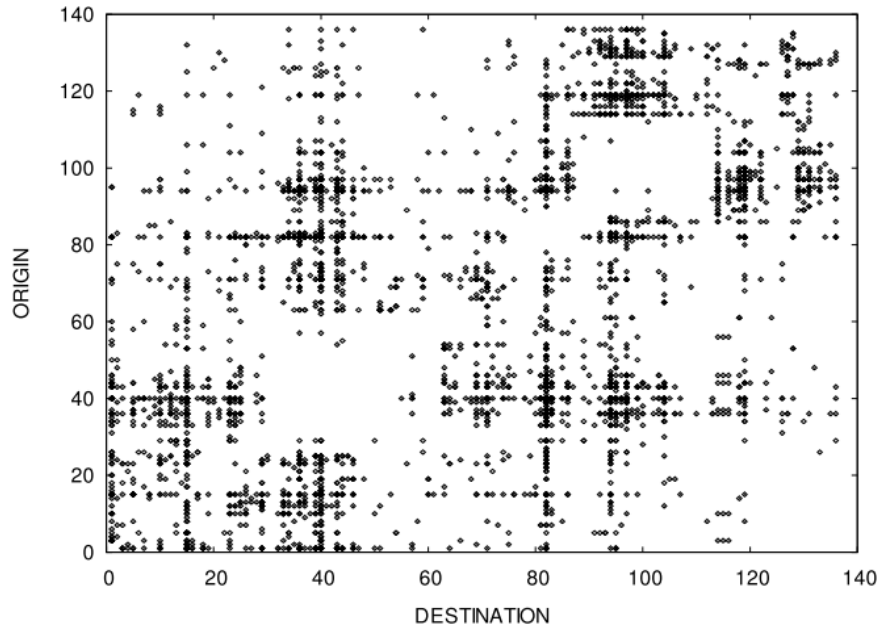
Figure 21. Spatial distribution, business trips by rail, Corridor OD matrix, 1719 non-zero flows
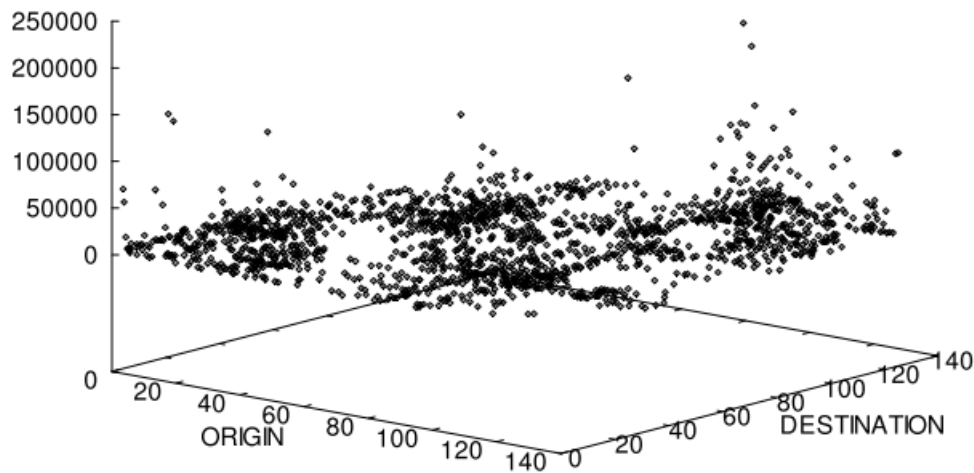
6.3. Other benefits from knowledge of curvatures

Although we have placed explicit emphasis above on Classical and Logit regression models, the issue of form arises with similar acuity for other types of regression analysis, such as Poisson and various count data methods, also used frequently in transportation, for instance to study accidents.

Fishy Poisson models. Some authors of Poisson models, such as Fridstrøm (2000a), have demonstrated the relevance of BCT use to the analysis of road accidents — even improving the Poisson methodology by allowing for the treatment of heteroskedasticity with type (6-D) specifications (Fridstrøm, 2000b) — but other authors avoid this progress in the establishment of

correlations. For instance, Dionne *et al.* (1992) surprisingly refrain from testing the linearity of the variables used in their initial Poisson model[120] of air accidents and insist on advancing for publication unchanged and untested form results (Dionne *et al.*, 1997) even claiming in particular that results for one of their variables, *Maintenance expenditures per aircraft departure*, were "very important" (p. 394) [121] despite six *t*-statistics averaging 1,71 in absolute value in their principal models (2, 3, 5 and 6). Would optimal form estimates have yielded really statistically significant coefficients, sign reversals, as in Table 13, or worse? The issue does arise in all count data models.

For our purposes, gross BCT curvatures on Fare and Time variables in Logit regression immediately matter in practice because there is little one can do about Distance attitudes and little about Risk attitudes. Best fit form values will deeply influence results for all other transformed variables as well, even if their immediate effect on forecasts is less dramatic.

Financial pitfalls of Linearity. In our HSR problem, the existence of linear model over-prediction and its exact amount make it insufficient to present results based on an untested assumption of linearity, or on a priori forms as currently done also for freight in some Corridor models (*e.g.* Patterson & *al.*, 2007)[122]. Generally, HSR project financing plans based on such assumptions are more than dubious because linear models tend to over-predict demand and revenues: form tests are not a fine tuning issue but a fundamental issue of scientific and practical relevance.

Economic pitfalls of Linearity. But this increased realism of unconditional correlation parameter estimates associated with demonstrably different quantity and revenue forecasts and market share distance profiles also increases the effort required for the determination of consumer surplus changes resulting from HSR investments, should these be of interest beyond financial quantities.

In practice, consumer surpluses are not calculated for compensated demand curves but for market demand curves, so the academic objection that the presence of non linearity will at least «perturb the *log-sum* formula» (McFadden, 1998) should not prove insurmountable. As, in any case, a Logit model is only a part of the full demand model, it is the evaluation of an integral of (13-B) over a certain range where total demand is endogenous that matters (Kikodoro, 2007). Linear Logit consumer surplus measures based on changes in the *log-sum* (Small & Rosen, 1981) will then have to take curvature into account, but must be completed by a calculation of surplus variations arising from changes in total travel estimated by often using, as Table 7 makes clear, something else than a *log* of the sum. Generally, the exact form will also be decisive in determining whether third degree monopolistic price discrimination can be welfare enhancing or not (Aguirre *et al.*, 2010).

We will not deal here with an important practical difficulty recently raised by Beuthe (2010) who studies problems that may occur in a careful analysis of the demand function over its full range when some BCT have powers greater than unity whereby shifts caused for instance by improved service quality, such as speed, formally imply infinite changes in consumer surplus, contrary to values smaller than unity for which the additional surfaces are bounded. This problem requires special treatment beyond the reach of this survey.

---

[120] In late 1991, the model was proposed to, and badly received by, the Canadian Royal Commission on National Passenger Transportation where it was felt that its untested linearity was inadequate to establish anything. The authors subsequently refused to consider the BCT form tests easily performed on explanatory variables of any Poisson model.

[121] In their preferred Model 2 of Table 2, the *t*-statistic of the intercept is (-5,12) and that of this variable is (-1,66); in the corresponding Model 5 of Table 3, the *t*-statistic of the intercept is (-2,35) and those of the same variable broken down between small and large firms are respectively (-2,46) for the small and (1,55) for the large firms. All other continuous variables, such as Hours flown and Speed, are also entered linearly and without form tests.

[122] The authors use a sample of 5670 observations from an SP shipper survey carried out in 2005. The Mixed Logit formulation apparently uses the *a priori* assumption that the product of the natural logarithm of cost and distance is normally distributed and the 3 alternatives appear to represent truck and the two competing railways. No justification is given for the log-normal assumption and other forms (*e.g.* square root-normal, *etc.*) are not considered.

7. The QDF Framework with Dogit and IPT-Logit cores

Defining building blocks: cores and representative utility functions. In this paper, a core consist in a set of modal utilities $U_i \geq 0$ appearing jointly in the Mode Split and in the Utility term U of the Total demand component of the QDF Modal demand architecture.

Up to this point, we have featured a Modal utility defined by the Logit quantity, but others are useful to define alternate families and will be explained shortly by giving the reader a sense of their origin and meaning before we document their use and relevance to functional form structures:

Table 17. Modal utilities or quantities for the QDF framework

Core	Modal utility or Quantity $U_m \geq 0$
Logit	$\exp(V_m)$
Standard Dogit	$\exp(V_m) + \theta_m \sum\limits_{j=1}^{M} \exp(V_j)$, $\theta_m \geq 0$
Generalized Dogit	$\exp(V_m) + \sum\limits_{j \neq m}^{M} \theta_{mj} \exp(V_j)$, $\theta_{mm} = 0$ and $\theta_{mj} \geq 0$
LIN-IPT-Logit	$\left[\phi_m \exp(V_m) + 1\right]^{1/\phi_m} - \mu_m$, $\phi_m \geq 0$ and $\mu_m \leq 1$
BT-IPT-Logit	$\exp\left\{ \dfrac{\left[\exp(V_m) + \tilde{\mu}_m\right]^{\tilde{\phi}_m} - 1}{\tilde{\phi}_m} \right\}$, $\tilde{\mu}_m \geq 0$

For any Quantity defined in Table 17, the representative utility function (RUF) proper can in principle obtain one of two specifications:

$$(35\text{-A}) \qquad V_i = \beta_{i0} + \sum_n \beta_{in}^i X_n^{(\lambda_{in}^i)} + \sum_{j \neq i} \sum_n \beta_{in}^j X_n^{(\lambda_{in}^j)} + \sum_s \beta_{is} X_s^{(\lambda_{is})} \qquad \text{(Generalized Box-Cox)}$$

$$(35\text{-B}) \qquad V_i = \beta_{i0} + \sum_n \beta_{in}^i X_n^i + \sum_{j \neq i} \sum_n \beta_{in}^j X_n^j + \sum_s \beta_{is} X_s \qquad \text{(Generalized Linear)}$$

and we have up to this point featured special cases, limiting ourselves[123] to

$$(35\text{-C}) \qquad V_i = \beta_{i0} + \sum_n \beta_{in}^i X_n^{(\lambda_{in}^i)} + \sum_s \beta_{is} X_s^{(\lambda_{is})} \qquad \text{(Standard Box-Cox)}$$

$$(35\text{-D}) \qquad V_i = \beta_{i0} + \sum_n \beta_{in}^i X_n^i + \sum_s \beta_{is} X_s \qquad \text{(Standard Linear)}$$

and to monotonic applications of the BCT to variables, thereby excluding on purpose

$$(36\text{-A}) \qquad y^{(\lambda_y)} = \beta_0 + \beta_1 X^{(\lambda_1)} + \beta_2 X^{(\lambda_2)} + \ldots$$
$$(36\text{-B}) \qquad V_i = \beta_0 + \beta_1 X^{(\lambda_1)} + \beta_2 X^{(\lambda_2)} + \ldots$$

(Turning effects)

analysed in detail in Gaudry *et al.* (2000) and sometimes applied outside of transport (*e.g.* Heckman & Polachek, 1974) or in transport with models of type (6-A), such as Tegnér *et al.* (2000), but

---

[123] Some are extremely rarely found, for instance (35-D). For a comparison of (35-B) and (35-D), see Laferrière & Gaudry (1992) where an excellent 1976 database of 211 morning peak observations on OD flow shares in Winnipeg made it possible to estimate the mode-specific constants for car and transit and their envelope curvatures within a LIN-IPT-L, starting from a previous specification developed by Cléroux *et al.* (1981).

unknown in those of type (6-B) where prevalence of linearity militates against them and dignified symmetric response curves could not conceivably go up and down, for instance with respect to Age.

Our intents in imposing these Quantity and RUF restrictions in previous sections were to put the emphasis on the potential asymmetry of Logit response curves because Logit cores dominate practice and because IIA-consistent representative utility functions suffice to make our case. But the new quantities enrich modelling possibilities in three ways that deserve examination as they might impact our stand:

(i) IPT-Logit quantities solve in principle the practical problem of identification of some parameters, in particular those of alternative-specific constants, met when Logit or Dogit quantities are used;

(ii) Dogit and IPT-Logit quantities allow for thick tails in response curves, an issue of modeller ignorance;

(iii) IPT-Logit quantities make it possible to search for non linearity not only of variables describing alternatives within RUF but also of Logit quantities, an issue of the allocation of non linearity present in response curves.

The identification of all alternative-specific constants. The identification problem arises because a reference alternative r must be found for the constant and the socioeconomic variable in

$$
(35\text{-D}) \quad
\begin{aligned}
V_i &= \left(\beta_{i0} - \beta_{r0}\right) + \sum_n \beta_{in}^i X_n + \sum_s \left(\beta_{is} X_s - \beta_{rs} X_s\right) \\
V_i &= \quad \beta_{i0}^{\nabla} \quad + \sum_n \beta_{in}^i X_n + \sum_s \quad \beta_{is}^{\nabla} \quad X_s
\end{aligned}
\qquad \text{(Standard Linear)}
$$

and only for the constant if alternative specific BCT are applied to the socioeconomic variable X_s.

In mode choice analysis, discussions have centered on how many of the identifiable *M-1* coefficients should be retained (Tardiff, 1978) and the problem has consistently nagged researchers (Bierlaire *et al.*, 1995). But the same problem seems intractable in path choice problems. If the application is binomial and the reference alternative becomes "any not chosen", as in the *tracé* choice application by McFadden (1968 or 1976a), the estimate effectively becomes one of differences between constants that are random variables (and estimators are *necessarily biased*) because there is no natural way to label alternative paths as there exist naturally meaningful ways to label modes. This predicament does not change if the application is multinomial and the identification of all coefficients of alternative specific constants becomes relevant as indicators of misspecification of RUF.

But use of IPT-L cores offer a solution, and the identification of all *M* coefficients of constants is made possible by IPT-L envelopes, as Laferrière (1988, 1999) demonstrated in practice for specific constants of air paths (defined sets of itineraries of identical Time and Fare characteristics) by OD pair in a Canada-wide model of Air demand built with a 100% sample (16 million individual trips) of domestic air trips made on Air Canada and Canadian Pacific Airlines in 1983.

The issue of asymptotes, thick tails, or modeller ignorance. Quite often, Logit cores fit badly at the ends of response curves. The first non-Logit core to address this problem of modeller ignorance, essentially determined by sparse RUF Logit specifications[124], is the Standard Dogit core (Gaudry & Dagenais, 1977 or 1979a). It adds θ_i parameters in the ith preference function and the Generalized

---

[124] A predominant form of RUF imperfection is incomplete lists of factors but there is no way to predict consequences for possible ecological aggregation error in share models and possible heterogeneous idiosyncrasy dominance in discrete choice models,

version transform these indices into θ_{ij} under certain conditions applicable to alternative specific constants but not to generic mode-abstract constants (Gaudry & Dagenais, 1981).

In both Dogit cases, the asymptotes of response curve tail ends will be modified but the above mode-specific coefficient under-identification problem will remain unsolved[125] and, due to the introduction of "relevant" cross-terms to own-mode utilities, neither model will generally be IIA consistent, although the issue is "dodged" because IIA-consistency could be present for some pair of alternatives if their θ parameters are zero. Dogit models therefore avoid IIA consistency without resorting to cross terms in LOS variables, the usual reason for holding IIA inconsistency as untenable (Samuelson, 1985).

Upon reception of the initial version of the Standard Dogit formulation (Gaudry & Dagenais, 1977, October), McFadden realized that he had formulated a few months before (McFadden, 1976b, December) an identical model which he had not made public because he felt that the log-likelihood function had lost its theoretical unimodality[126], already noted above as an obstacle to the use of non linear forms. Ben-Akiva, aware of McFadden's, then wrote another document (Ben-Akiva, 1977 November) later diffused in "substantially modified version" (Ben-Akiva & Swait, 1984) and later with a modified title (Ben-Akiva & Swait, 1987) where he called the Dogit model the "parametrized captivity model", recently shortened to "Logit captivity model" (Walker & Ben-Akiva, 2011).

If a new name had been given to every new derivation of the Gravity model or of the Logit model, taxonomic confusion would reign. Fortunately, the "compressed/saturated binary Logit" for rail (Westley, 1979) was rapidly shown to be a special case of Dogit (Hensher, 1982) and there was no further confusion between derivation and name when the Dogit was applied to OD pairs instead of modes (Chu, 1990), or more generally (Brown, 1980; Chu, 2011), and derived as a "brand loyalty" model even if buyers are not perfectly captive to one brand (Bordley, 1989, 1990) and more generally (Sonis, 1992)[127], or when its statistical properties were studied by Williams & Ortúzar (1982) among others (Fry, 1988; Fry & Harris, 1995).

This captivity derivation and interpretation of the Standard Dogit has some appeal because the simplest way to understand this model is to consider that one is drawing from a mixed distribution (McFadden, 1981) with two populations: one captive to any alternative in proportion C_i and the remaining non-captive and exercising a choice according to the Logit model, as in:

(37)
$$P_i = \frac{\theta_i}{(1 + \sum_j \theta_j)} + \frac{1}{(1 + \sum_j \theta_j)} \bullet \frac{\exp(V_i)}{\sum_j \exp(V_j)} = \frac{\exp(V_i) + \theta_i \sum_j \exp(V_j)}{(1 + \sum_j \theta_j) \sum_j \exp(V_j)}$$

$$P_i = C_i + (1 - \sum_j C_j) \bullet \left[\frac{\exp(V_i)}{\sum_j \exp(V_j)} \right]$$

This interpretation of the θ_i parameters parallels the interpretation of the "basic consumption" parameters found in Stone's (1954) linear expenditure model and leads to the same ambiguities.

---

[125] And some normalization, based on choosing a reference alternative, is required. In Table 17, the Generalized Dogit is written with a normalization rule that makes setting all non diagonal values of θ_{mj} to zero directly yield the Logit and equal by row the Standard Dogit.

[126] Conversation in Marc Gaudry's office, before Daniel McFadden's October 26, 1977 *Université de Montréal* seminar. Since then, multiple maxima have not yet been found in practice, either with aggregate or with discrete data.

[127] Sonis' starting point was that the multiplication of some specifically chosen Markovian matrices on the vector of choice probabilities (used to explain innovation diffusion and dynamic choice) for teh Logit model gives the vector of choice probabilities for the Standard Dogit model.

Indeed, in a number of empirical studies of linear expenditure models, Solari (1971) has found that the best fits were obtained when certain of these parameters were negative[128]. He therefore reinterpreted such negative values as indicative of "superior" goods. In our strictly comparable situation, are we to say that negative θ_i values are indicative of a *free* population, positive ones of a *captive* population and the rest of a *normal* population? In the same vein, a limited number of experiments involving four intercity modes in Canada (Dagenais *et al.*, 1982a, 1982b) suggested that better fits were obtained when the θ_i parameters were not constrained to be positive.

In Table 17, we therefore constrain Dogit parameters to be positive, avoid making sense of negative captivity and negative basic needs, and adopt the "modeller ignorance" interpretation consistent with incomplete regressor lists in RUF presumed to contain the researcher's "knowledge" or information[129]. The non negativity constraint also guarantees non negative Modal utilities U_m, as do other constraints in Table 17. We illustrate such flexible asymptotic non negative ignorance limits $\theta_i \geq 0$ in Figure 22 with a case where Linear Logit symmetry is maintained between the new limits.

Figure 22. Linear-Logit *vs* Standard-Dogit
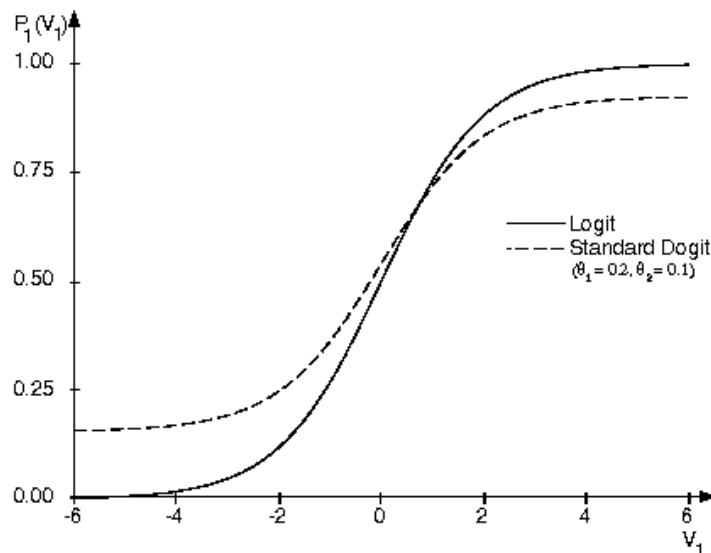
The Generalized version (Gaudry & Dagenais, 1981) is close in spirit to Restle (1961) and akin to a transactions cost approach where the relevance of other alternatives depends on some measure of similarity/dissimilarity (to use Restle's language) implemented for instance by Cascetta *et al.* (1996) who introduce in the RUF of an alternative an average value for other alternatives.

Combining modeller ignorance and allocation of non linearity. The second non Logit core to contain flexible modeller ignorance limits is the IPT approach (Gaudry, 1978 or 1981), where use is made of Tukey's μ location parameter (Anscombe & Tukey (1954) referred to in Tukey, 1957) used in (6-C). In the direct BT garb of (6-C), this μ applied to variables is rarely successful except as a geometric device to avoid natural coordinates and make them endogenous, or as a diagnostic tool (Atkinson, 1983). But in the indirect BTG garb of (6-C), the same μ applied to functions acquires further meaning as an ignorance parameter. Applied to Logit quantities, an IPT-Logit core permits to allocate non linearity elsewhere than to variables of RUF functions. Despite a successful

---

[128] During his *Université de Montréal* visit around 1975, he quipped that he had never seen a sufficiently disaggregated model yielding only positive "basic need" indicators.

[129] In a ten-variable 5-mode urban model for São Paulo, it is found that, if three of the modes are assigned θ_i, the log-likelihood only gains 3,6 points (Swait & Ben-Akiva, 1987). Clearly, the more information is provided, the weaker the presence of modeller ignorance due to a better explanation of tails.

application outside of transportation of the LIN-IPT-Logit[130] core (Montmarquette & Mahseredjian, 1985), IPT-Dogit or IPT-Logit cores have never been derived. We consider IPT-Logit cores here and neglect IPT-Dogit cores. The illustration in Figure 23 modifies the symmetry and the asymptotic limits of the Logit core specification[131].

Figure 23. Linear-Logit *vs* Box-Tukey-Inverse-Power-Transformation-Logit
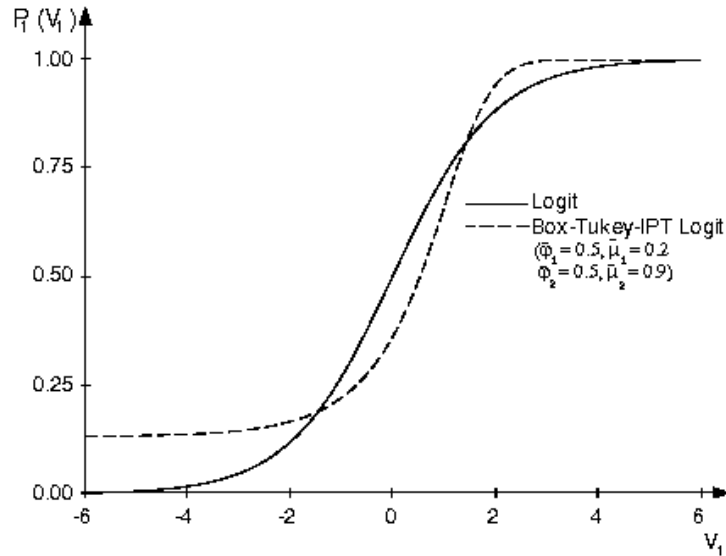
Do enriched core results confirm those of Tables 7-9? The question is whether enriched cores all containing the Logit core as a nested special case should change the analysis, made in Tables 7, 8 and 9, of LOS non linearity estimated under certainty assumptions. To answer it, consider Table 18.

Table 18. BCT estimates for Time & Cost variables in passenger models with non Logit cores

Column	1	2	3	4	5	6	7	8	9	10	11	
Core	Logit			Standard Dogit			BT-IPT-Logit			*Envelope*	*Ignorance*	
	λ_T	λ_F	$\lambda_T - \lambda_F$	λ_T	λ_F	$\lambda_T - \lambda_F$	λ_T	λ_F	$\lambda_T - \lambda_F$	φ	μ	**Source**
Intercity												
44. Canada 1976 (2 m)	0,36	0,79	-0,42	0,23	0,70	-0,47				n.a.		Table 2, Col. DU and LU
Log-likelihood		**133,31**			**134,52**					n.a.	Weak	**Gaudry & Wills, 1979**
45. France + Channel (3 m)	-0,62	-0,62	0,00				-0,42	-0,42	0,00			Table 10, Col. 1.2 and 1.5
Log-likelihood		**-6246,86**						**-5481,88**		Strong	Strong	**Gaudry *et al*. 1998**
46. Canada 1976 (4 m)	-0,01	-0,01	0,00				1,00	1,00	0,00			Table 4, Col. 2 and 6
Log-likelihood		**-499,345**						**-490,47**		Moderate	Moderate	**Gaudry, 1990 or 1993**
Urban												
47. Montreal 1970 (2 m)	0,66	4,21	-3,56	0,56	4,50	-3,94				n.a.		Table 1, Col. U and U-U
Log-likelihood		**202,97**			**203,32**					n.a.	Weak	**Gaudry, 1980**
48. Santiago 1983-1985 (9 m)	1,63	0,35	1,28	1,84	0,89	0,95	2,29	1,27	1,01			Col. 3, 4, 7
Log-likelihood		**-837,93**			**-815,92**			**-817,82**		Strong	Strong	**Tran & Gaudry, 2009b**

Top models beyond 40. We assume, and will demonstrate shortly in Table 19, that enriched cores do not modify the meaning of BCT parameters used on LOS variables.

The analysis of Table 18 results therefore show that: (i) core enrichments do not significantly change the estimates obtained with a Logit core except for Model 48 (Santiago de Chile) where relative damping fluctuates in the range of slope independence $(\lambda_T - \lambda_F) = 1,00$; (ii) there is more

---

[130] It includes $f(x) = \exp\left[\exp(x)\right]$ related to the so called (Johnson & Kotz, 1970, p. 273) log-Weibull $f(x) = \exp\left[-\exp(-x)\right]$ if $\phi_m = 0$ and $\mu_m = 0$.

[131] Figures 20 and 21 are generated with the same Model 0 (see Table 12) used for Figure 1 and specified in its vicinity.

Table 19. QDF demand elasticities and slopes for five cores (variable X in RUF$_m$)

(38)	**Logit**
η_m	$\beta_U \dfrac{U^{\lambda_U}}{T_{TOT}^{\lambda_{TOT}}} \beta_X X^{\lambda_{mX}} P_m + \beta_X X^{\lambda_{mX}} (1 - P_m)$
δ_m	$\beta_X X^{\lambda_{mX}-1} T_m \left[\beta_U \dfrac{U^{\lambda_U}}{T_{TOT}^{\lambda_{TOT}}} P_m + (1 - P_m) \right]$

(39)	**Standard Dogit**
η_m	$\beta_U \dfrac{U^{\lambda_U}}{T_{TOT}^{\lambda_{TOT}}} \dfrac{\beta_X X^{\lambda_{mX}} \exp(V_m)(1 + \sum\limits_{m=1}^{M} \theta_m)}{\sum\limits_{m=1}^{M}\left[\exp(V_m) + \theta_m \sum\limits_{j=1}^{M} \exp(V_j) \right]} + \dfrac{\beta_X X^{\lambda_{mX}} (1 - P_m)(1 + \theta_m)}{1 + \theta_m \sum\limits_{j=1}^{M} \exp(V_j - V_m)}$
δ_m	$\beta_X X^{\lambda_{mX}-1} T_m \left[\beta_U \dfrac{U^{\lambda_U}}{T_{TOT}^{\lambda_{TOT}}} \dfrac{\exp(V_m)(1 + \sum\limits_{m=1}^{M} \theta_m)}{\sum\limits_{m=1}^{M}\left[\exp(V_m) + \theta_m \sum\limits_{j=1}^{M} \exp(V_j) \right]} + \dfrac{(1 - P_m)(1 + \theta_m)}{1 + \theta_m \sum\limits_{j=1}^{M} \exp(V_j - V_m)} \right]$

(40)	**Generalized Dogit**
η_m	$\beta_U \dfrac{U^{\lambda_U}}{T_{TOT}^{\lambda_{TOT}}} \dfrac{\beta_X X^{\lambda_{mX}} \exp(V_m)(1 + \sum\limits_{j \neq m}^{M} \theta_{mj})}{\sum\limits_{m=1}^{M}\left[\exp(V_m) + \sum\limits_{j=m}^{M} \theta_{mj} \exp(V_j) \right]} + \dfrac{\beta_X X^{\lambda_{mX}} (1 - P_m)}{1 + \sum\limits_{j \neq m}^{M} \theta_{mj} \exp(V_j - V_m)}$
δ_m	$\beta_X X^{\lambda_{mX}-1} T_m \left[\beta_U \dfrac{U^{\lambda_U}}{T_{TOT}^{\lambda_{TOT}}} \dfrac{\exp(V_m)(1 + \sum\limits_{j \neq m}^{M} \theta_{mj})}{\sum\limits_{m=1}^{M}\left[\exp(V_m) + \sum\limits_{j=m}^{M} \theta_{mj} \exp(V_j) \right]} + \dfrac{(1 - P_m)}{1 + \sum\limits_{j \neq m}^{M} \theta_{mj} \exp(V_j - V_m)} \right]$

(41)	**Linear Inverse Power Transformation-Logit**
η_m	$\beta_U \dfrac{U^{\lambda_U}}{T_{TOT}^{\lambda_{TOT}}} \beta_X X^{\lambda_{mX}} \left[\exp(V_m) + \tilde{\mu}_m \right]^{\tilde{\phi}_m - 1} P_m \dfrac{\left[\phi_m \exp(V_m) + 1 \right]^{\frac{1}{\phi_m} - 1} \exp(V_m) \beta_X X^{\lambda_{mX}} (1 - P_m)}{\left[\phi_m \exp(V_m) + 1 \right]^{1/\phi_m} - \mu_m}$
δ_m	$\beta_X X^{\lambda_{mX}-1} T_m \left[\beta_U \dfrac{U^{\lambda_U}}{T_{TOT}^{\lambda_{TOT}}} \left[\exp(V_m) + \tilde{\mu}_m \right]^{\tilde{\phi}_{rail} - 1} P_m + \dfrac{\left[\phi_m \exp(V_m) + 1 \right]^{\frac{1}{\phi_m} - 1} \exp(V_m)(1 - P_m)}{\left[\phi_m \exp(V_m) + 1 \right]^{1/\phi_m} - \mu_m} \right]$

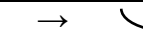| (42)	**Box-Tukey Inverse Power Transformation-Logit**
η_m	$\beta_U \dfrac{U^{\lambda_U}}{T_{TOT}^{\lambda_{TOT}}} \dfrac{\beta_X X^{\lambda_{mX}} \left[\exp(V_m) + \tilde{\mu}_m \right]^{\tilde{\phi}_{rail} - 1} \exp\left\{ \dfrac{\left[\exp(V_m) + \tilde{\mu}_m \right]^{\tilde{\phi}_m} - 1}{\tilde{\phi}_m} \right\}}{\sum\limits_{m=1}^{M} \exp\left\{ \dfrac{\left[\exp(V_m) + \tilde{\mu}_m \right]^{\tilde{\phi}_m} - 1}{\tilde{\phi}_m} \right\}} + \left[\exp(V_m) + \tilde{\mu}_m \right]^{\tilde{\phi}_m - 1} \exp(V_m) \beta_X X^{\lambda_{mX}} (1 - P_m)$
δ_m	$\beta_X X^{\lambda_{mX}-1} T_m \left[\beta_U \dfrac{U^{\lambda_U}}{T_{TOT}^{\lambda_{TOT}}} \dfrac{\left[\exp(V_m) + \tilde{\mu}_m \right]^{\tilde{\phi}_m - 1} \exp\left\{ \dfrac{\left[\exp(V_m) + \tilde{\mu}_m \right]^{\tilde{\phi}_m} - 1}{\tilde{\phi}_{rail}} \right\}}{\sum\limits_{m=1}^{M} \exp\left\{ \dfrac{\left[\exp(V_m) + \tilde{\mu}_m \right]^{\tilde{\phi}_m} - 1}{\tilde{\phi}_m} \right\}} + \left[\exp(V_m) + \tilde{\mu}_m \right]^{\tilde{\phi}_{rail} - 1} \exp(V_m)(1 - P_m) \right]$
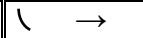
absolute amplification (see braided cases)[132] than in Tables 7 to 9; (iii) Model 47 for trips (all purposes) in the Montreal area is the first among 11 comparable (two LOS BCT) urban models not to show absolute Time amplification.

We may note in Table 21 of Appendix B how the η_m expressions of Table 19 are affected when a LOS variable[133] of interest X appears in more than one RUF in accordance with (35-A), rather than on the lines of (35-C) assumed in Table 19.

Making sense of curvatures. We now have to confirm that more complex specifications of cores do not change our previous interpretation of BCT LOS values. For this, we first recall in Table 19 the Logit core elasticities η_m and demand slopes δ_m from (13-A)-(13-B) and, after a simplification of indices to make the expressions apply to any mode m, compare them to corresponding expressions obtained with the four other enriched cores.

One can immediately see that former definitions of damping, independence and amplification ─ whether absolute in (13-D) to (13-F) or relative in (15-C) to (15-E) ─ follow through, as does the definition of the VOT, either without Distance explicitation in (14-A) or with Distance explicitation in (14-B). Together, this gives for memory Table 20.

Table 20. Absolute and relative effects defined by values of LOS BCT or by their differences

	λ_T or λ_F			$\lambda_T - \lambda_F$		
Effect of ΔLOS on the:	<1	=1	>1	>0	=0	<0
	$\searrow \;\rightarrow\; \searrow$	$\backslash \;\rightarrow\; \backslash$	$\searrow \;\rightarrow\; \frown$	▲	=	▼
	damped	*none*	*amplified*	*damped*	*none*	*amplified*
	slope of demand curve [from (13-D) to (13-F)]			value of Time [from (15-C) to (15-E)]		

---

[132] Model 44 aggregates the three public modes by a procedure based on averages superseded since by the log-sum aggregator. If no aggregation is carried out with the same data, as in Model 1 of Table 7, amplification disappears. Among all 48 models listed in Tables 7-9 and 18, Model 44 it is the only one to use aggregates of modes.
[133] The derivations of the elasticity values are found in Tran & Gaudry (2010a).

8. Conclusion

We have argued that, as envisaged High Speed Rail (HSR) projects typically anticipate a reduction by half or more of existing rail Travel Time, ridership and revenue forecasts depend critically on the curvature of the rail demand curve defined with respect to both Fare and Time levels of service (LOS), in addition to Frequency. We have discussed these curvature issues and derived forecasts within the Quasi-Direct Format (QDF) architecture in use since the US Northeast Corridor Transportation Project of the 1960's, a product of Total Market size and Mode split components.

We have studied curvatures with Box-Cox transformations (BCT) applied to the variables of both component models but our discussion has been focussed on the Logit piece used in representative QDF structures because modal diversion effects of new HSR services tend to dominate the slower long-term induction effects of increased market size, so that net discounted project results are driven by Linear Logit specifications prevailing to this day in transport, but also elsewhere.

Our analysis started with a reference summary of results from three Canadian mode choice models formulated in 1976-1978 and in 1992-1994 for the purposes of forecasting the effects of major infrastructure changes, respectively new airports in Southern Ontario and faster rail in the Quebec-Windsor Corridor. We have emphasized how HSR revenue maximization of rail Fares and Speeds obtained under hypothesized Linear forms then yielded higher revenues than under data determined optimal Box-Cox forms even if their market shares were higher on relatively long distances. The rest of the paper interpreted or positioned these results among others and further documented Corridor model forecasts.

Concerning their interpretation, we have pointed out that, although basic consumer demand theory does not constrain admissible values of BCT in Total demand and Logit Mode choice components of modal demand models, actual estimates are in fact generally compatible with "Cost damping" claims: (a) Time and Cost sensitivities (expressed as first partial derivatives of the demand function) typically fall with Distance in passenger and freight markets, except in urban passenger markets where Time sensitivity almost always increases; (ii) relative sensitivities (the Value of Time) always increase with Distance, irrespective of whether the absolute value of slopes falls at a decreasing (damped) or at an increasing (amplified) rate, in the 40-some surveyed models built by some 30 researchers for 10 countries. Such empirical regularities establish a fundamental difference between intercity and urban markets based on the second derivatives of demand functions.

We have further suggested that such real gross BCT power value sensitivity profiles estimated without taking the attitude to risk into account could in fact reflect two effects that, as shown in recent seminal work on Rank Dependent Utility (RDU), can be identified by products of power functions of Fare or Time, to wit a simple power to determine *the attitude to outcome risk (probability)* and a BCT power to determine the *attitude to outcome proper*. We have also reinterpreted recent models making successful use of interactions between a Distance variable raised to a simple power and a LOS variable (especially Travel Time) sometimes raised to a BCT power as identifying, also by a product of functions, an "attitude to Distance" that allows for an explicit breakdown of gross attitudinal parameters between *attitude to distance* and *attitude to outcome* elements.

We have implied overall that, beyond mere fit and other demonstrated benefits, *untested* linear forms of Standard Logit utility function variables are theoretically unexpected as representations of price-time utility maps, statistically unsustainable in many samples ranges where *prima facie* gross cost or time damping or amplification prevail in absolute and relative senses, practically biased as conditional bargaining games to play with data, and often demonstrably unsound or misleading in the production of HSR passenger and revenue forecasts, and no doubt elsewhere as well.

9. Appendix A. On estimating simple powers instead of BCT

We document here the statement, made in this main body of the text, that the QDF structure (7-A) is preferable to the CES Armington structure in part because the latter, based on simple power transformations, notably presents a number of estimation difficulties which do not arise with BCT. We explain that the safe way to obtain CES power values is to estimate BCT transformations, first and foremost in order to be certain of obtaining power estimates that maintain the order of the data.

To start, it is obvious that maximization of the Log likelihood of the dependent variable in models with explanatory terms such as (6-G), and more generally models of form (7-B), admits of degenerate solutions if powers like β_U or σ are set at zero: to avoid such model collapse, users of (6-G) typically force the elasticity of substitution σ to be greater than 1, thereby excluding a domain of values potentially consistent with the data. But another, more serious, problem lurks even with constrained simple power values because, in addition to lacking important continuity at zero, simple powers do not generally preserve the order of the data. We illustrate the ordering of the data first and demonstrate afterwards the continuity at zero assumed in (6-C) above.

Consider Figure 24 from Johnston (1984, p. 63), where two data values, 10 and e = 2,84128, are transformed in three ways: in (a) by a simple power transformation y^λ, which leads to an inversion of values at point (0,1); in (b) by y^λ/λ, which maintains the order but causes a discontinuity at $\lambda = 0$; in (c) by $(y^\lambda - 1)/\lambda$, which preserves both continuity at zero and the order of data values[134]. Basically, all Constant Elasticity of Substitution (C.E.S.) models pose these estimation problems.

Figure 24. Illustration of the continuity at 0 and of the order preservation of the BCT
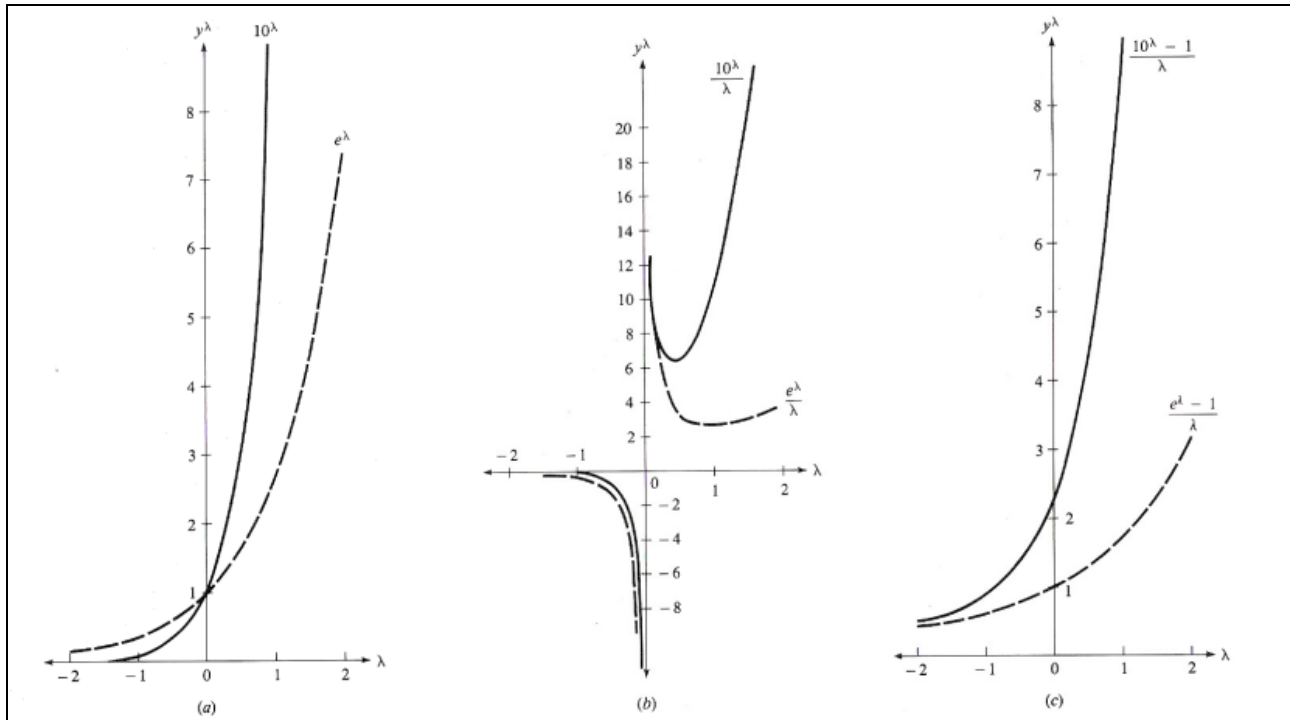
Concerning continuity at zero, we want to prove that the common Box-Cox transformation of a strictly positive variable y is continuous at 0 and is more precisely equal to the natural logarithm of y, as stated in (6-C):

---

[134] As $10 > e$, the order is modified in (a) at the point where λ changes sign, because $10^\lambda < e^\lambda$ if $\lambda < 0$ and $10^\lambda > e^\lambda$ if $\lambda > 0$; but it is preserved in (c) because $[(10^\lambda - 1)/\lambda] > [(e^\lambda - 1)/\lambda]$ for any value of λ.

(35) $\quad \dfrac{y^{\lambda}-1}{\lambda} \;\to\; \ln y \quad$ if $\lambda \to 0,$

a result that we derive in two ways:

    A. By application of L'Hospital's rule, namely

(36) $\quad \lim\limits_{\lambda \to 0}\left[\dfrac{y^{\lambda}-1}{\lambda}\right] = \lim\limits_{\lambda \to 0}\left[\dfrac{\frac{\partial (y^{\lambda}-1)}{\partial \lambda}}{\frac{\partial \lambda}{\partial \lambda}}\right] = \lim\limits_{\lambda \to 0}\left[\dfrac{y^{\lambda}\ln y}{1}\right] = \ln y$

    B. By a Taylor expansion of $f(\lambda) = y^{\lambda}$, $y > 0$ and $-\infty < \lambda < \infty$, around the point $\lambda = 0$:

(37-A) $\quad f(\lambda) = f(0) + \lambda f'(0) + \dfrac{\lambda^2}{2!}f''(0) + \dfrac{\lambda^3}{3!}f'''(0) + \ldots$

    Indeed, given that here

$$f'(\lambda) = y^{\lambda}\ln y \quad \Rightarrow f'(0) = \ln y$$
$$f''(\lambda) = y^{\lambda}\left[\ln y\right]^2 \quad \Rightarrow f''(0) = \left[\ln y\right]^2,$$
$$f'''(\lambda) = y^{\lambda}\left[\ln y\right]^3 \quad \Rightarrow f'''(0) = \left[\ln y\right]^3$$

    we can rewrite (37-A) as:

(37-B) $\quad f(\lambda) = 1 + \lambda\left[\ln y\right] + \dfrac{\lambda^2}{2!}\left[\ln y\right]^2 + \dfrac{\lambda^3}{3!}\left[\ln y\right]^3 + \ldots$

(37-C) $\quad \dfrac{f(\lambda)-1}{\lambda} = \left[\ln y\right] + \dfrac{\lambda^2}{2!}\left[\ln y\right]^2 + \dfrac{\lambda^3}{3!}\left[\ln y\right]^3 + \ldots$

(37-D) $\quad \dfrac{y^{\lambda}-1}{\lambda} = \left[\ln y\right]\left\{1 + \dfrac{\lambda}{2!}\left[\ln y\right] + \dfrac{\lambda^2}{3!}\left[\ln y\right]^2 + \ldots\right\}$

    the last form of which, if $\lambda = 0$, effectively reduces to:

(37-E) $\quad \dfrac{y^{\lambda}-1}{\lambda}\bigg|_{\lambda=0} = \ln y$

The Box-Cox transformation used in the QDF framework (7-A) avoids all three critical pitfalls: degeneracy, discontinuity at zero and data order perturbation. It is therefore unwise to estimate CES models, be they production or demand functions, as simple power models unless great care is taken. In that sense, BCT estimation is the correct way to estimate all simple power models, and notably CES specifications; it is more "idiot proof", lest one prefer risk or have no choice because, as in RDU models of type (17-A), a BCT choice for the function $\psi : [0,1] \to [0,1]$ would not maintain the "capacity" of the distribution, namely the property that the sum of probabilities be equal to one.

10. Appendix B. On the Transport, Hicksian and Slutsky decompositions

In this appendix, we first show how the transport and classical economic decompositions can be presented, at least for some special cases, on the same graph, but we do not work out the structure of indifference maps required, for each decomposition, to contrast the visual implications of constant marginal utility and more general indifference curve assumptions.

This superposition of decompositions is made under the assumption that type (5) Modal utility functions contain only own-mode characteristics. But, in a second step, we point out, using utility functions of type (8-A) where cross-modes may be present, that the sign of the transport Diversion effect can change when non separable utility is allowed and the Mode split model structure brought back into the traditional fold of sensible microeconomic demand system postulates for close substitutes.

A 45° final budget line simplification. Assume in Figure 25 that combinations of goods _**1**_ and _**2**_ are utility maximizing bundles and that, following a lowering of the price of mode 1, the budget line passing through combination _**2**_ is at a 45° angle. This assumption makes it possible to superpose parrallel lines to this final budget line, for instance a parallel that can be made to go through point _**1**_, and a constant Total trip line like T_1 or T_2 from Figure 6.

Figure 25. Three decompositions of the variation in Total demand from point 1 to point 2
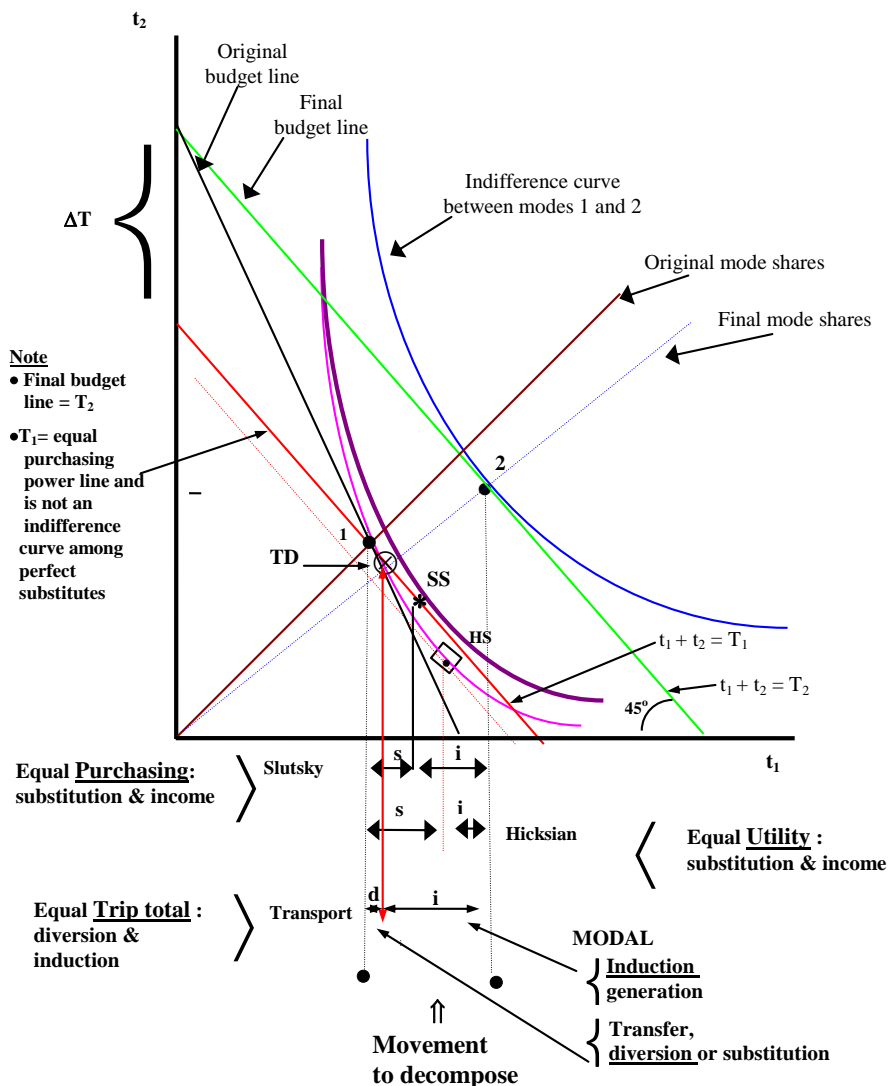
This superposition is tantamount to assuming the visual identity of Slutsky's *equal budget* line to a specific *equal total trip* (equal total output) line of transport models: it makes it possible to compare on the same graph the three decompositions of interest.

Classical micro-economic decompositions. To define Slutsky's substitution effect, one needs to find the point of maximum utility *SS* on the hypothetical budget line which reflects the new relative prices prevailing at point *2* but that makes it possible to purchase former combination *1*. Compared to point *1*, point *SS* is not at equal output but at **equal budget** (*purchasing power*). The movements $\{[1 \rightarrow SS] ; [SS \rightarrow 2]\}$ are the substitution and income effects that result from Slutsky's decomposition.

To define Hicks' substitution effect, one needs to find the tangency point *HS* between the initial indifference curve that determines the choice of point *1* and the minimum budget line incorporating the prices prevailing at point *2*. Compared to point *1*, point *HS* is not at equal budget or output but at **equal utility**. The movements $\{[1 \rightarrow HS] ; [HS \rightarrow 2]\}$ are the substitution and income effects that result from Hicks' decomposition.

We have conjectured in the main body of the paper that the ratio of Induction to Diversion that obtains in (16), being independent from prices (and from Income), should have analogous consequences for the ratios of Income to Substitution effects defined by Slutsky and by Hicks, but we shall not attempt to work this out here as it would require going in depth into the shape of indifference curves, an issue already partly explored before in Bruzelius (1979, Ch. 3).

Suffice it for our purposes to point out the most important differences among the three groups of effects: (i) the Induction effect is necessarily positive, which need not be the case for the Income effects it resembles, even if we expect them to be usually so; (ii) Slutsky's and Hicks' Substitution effects are necessarily positive, which need not be the case for the Diversion effect in general even if we expect it to be so (as shown in Figure 6). We now turn to this unexpected feature of Diversion.

On the sign of the Diversion effect. If the double IIA blinkers associated with Origin-Destination and Mode indices of the QDF structure used above are removed, the sign of the Diversion effect can indeed change. To show this, we recall from (1)-(3) in Section 2 that T_{ijm}, the Demand for a particular mode *m* from *i* to *j*, is obtained as a product of a model of the Total transport market by all modes T_{ijTOT} and a model of Modal split P_{ijm}. Neglecting *ij* subscripts, this can be restated for convenience as

$$(43\text{-}A) \qquad T_m = \{T_{TOT}\} \bullet \{P_m\},$$

or, more explicitly, as:

$$(43\text{-}B) \qquad T_m = \left\{ f(A_c, A_d, U) \right\} \bullet \left\{ U_m / \sum_m U_m \right\}, \quad m = 1,...,M,$$

where the model of Total demand by all modes contains vectors of activity variables A_c and A_d, such as Population and Income, and an index *U* of the utility of travel by all modes, called the coupling term or *U* term, was defined by the denominator of the mode split model:

$$(43\text{-}C) \qquad U = \sum_m U_m, \ U_m \geq 0,$$

where each Modal utility U_m term summarizes the attractiveness of a mode. This framework admits of many Mode split models, either aggregate (explaining market shares) or discrete (explaining categorical individual choices) but we have concentrated above on IIA features within the Mode choice component, only own-OD terms being considered (no cross-OD terms).

Because (43-A) is a product of models, the elasticity of the Modal demand (the left-hand-side variable), can be expressed as a sum of their elasticities:

(43-D)   [η of Demand for Mode m] = [η of Total Demand for all modes]+[η of Modal Split of Mode m],

but it is important to note that, in coupled products of paired models, a given explanatory variable X_k, such as consumer Income or the Price of a mode, might well appear in all three terms (A), (C) and (D) distinguishable by rewriting (43-D) in a more general and explicit fashion as:

(43-E)

$$\eta(T_m, X_k) = \underbrace{\tilde{\eta}(T_{TOT}, X_k) + \underbrace{\eta(T_{TOT}, U) \bullet \eta(U, X_k)}_{\text{(B)} \quad \text{(C)}}}_{} + \eta(P_m, X_k)$$

$$(F) \quad = \qquad (E) \qquad \qquad (D)$$

(A) (B) (C) ,

two particular cases of which can be considered:

a. If X_k appears only in the subset of X_k determining Total demand, but not in U, then $\eta(U, X_k)$ and $\eta(P_m, X_k)$ are equal to zero and the Modal and Total elasticities are identical:

(43-F) $$\eta(T_m, X_k) = \tilde{\eta}(T_{TOT}, X_k) = \eta(T_{TOT}, X_k)$$

b. Alternately, if X_k appears only in U, but not in the subset of X_k determining Total demand, then $\tilde{\eta}(T_{TOT}, X_k)$ disappears and the modal elasticity reduces to:

(43-G) $$\eta(T_m, X_k) = \eta(T_{TOT}, U) \bullet \eta(U, X_k) + \eta(P_m, X_k),$$

selected above to write demand elasticity formula (13-A) and study its embedded slope (13-B).

For the general structure (43-E), the sign of the modal elasticity $\eta(T_m, X_k)$ calculated as (F) will depend on the signs of the four elasticities $\tilde{\eta}(T_{TOT}, X_k)$, $\eta(T_{TOT}, U)$, $\eta(U, X_k)$ and $\eta(P_m, X_k)$. Since Total demand T_{TOT} is an increasing function of the utility index U, the elasticity $\eta(T_{TOT}, U)$ is necessarily positive. Assuming $\tilde{\eta}(T_{TOT}, X_k)$ has the same sign as $\eta(U, X_k)$, the sign of the Modal elasticity will be given by a combination of signs of the two elasticities $\eta(U, X_k)$ and $\eta(P_m, X_k)$ and two cases can be considered:

c. If X_k appears only in mode m, but not in other modes, then $\eta(U, X_k)$ and $\eta(P_m, X_k)$ have always the same sign:

(43-H) $$\eta(P_m, X_k) = \frac{\partial P_m}{\partial X_k} \frac{X_k}{P_m} = \frac{1}{U}(1 - P_m)\frac{\partial U_m}{\partial X_k}\frac{X_k}{P_m} = (1 - P_m)\frac{\partial U_m}{\partial X_k}\frac{X_k}{U_m},$$

d. If X_k appears in mode m and in at least one of the other modes, then $\eta(U, X_k)$ and $\eta(P_m, X_k)$ can have the same or opposite signs, as should be clear from:

(43-I)

$$\eta(P_m, X_k) = \frac{\partial P_m}{\partial X_k} \frac{X_k}{P_m} = \frac{1}{U}\left[\frac{\partial U_m}{\partial X_k} - P_m \frac{\partial U}{\partial X_k}\right]\frac{X_k}{P_m}$$

$$= \frac{\partial U_m}{\partial X_k}\frac{X_k}{U_m} - \frac{\partial U}{\partial X_k}\frac{X_k}{U} = \eta(U_m, X_k) - \eta(U, X_k),$$

$$= \left(\frac{1 - P_m}{P_m}\right)\eta(U, X_k) - \sum_{j \neq m}\left(\frac{P_j}{P_m}\right)\eta(U_j, X_k)$$

In this case, due to the presence of a second term which is a weighted sum of all cross elasticities, the signs of $\eta(U, X_k)$ and $\eta(P_m, X_k)$ clearly need not be the same: the sign of the

Modal elasticity does not necessarily follow the sign of $\eta(U, X_k)$ in such cases, but this is difficult to detect from Table 21[135] except by imagining higher cross effects than own effects.
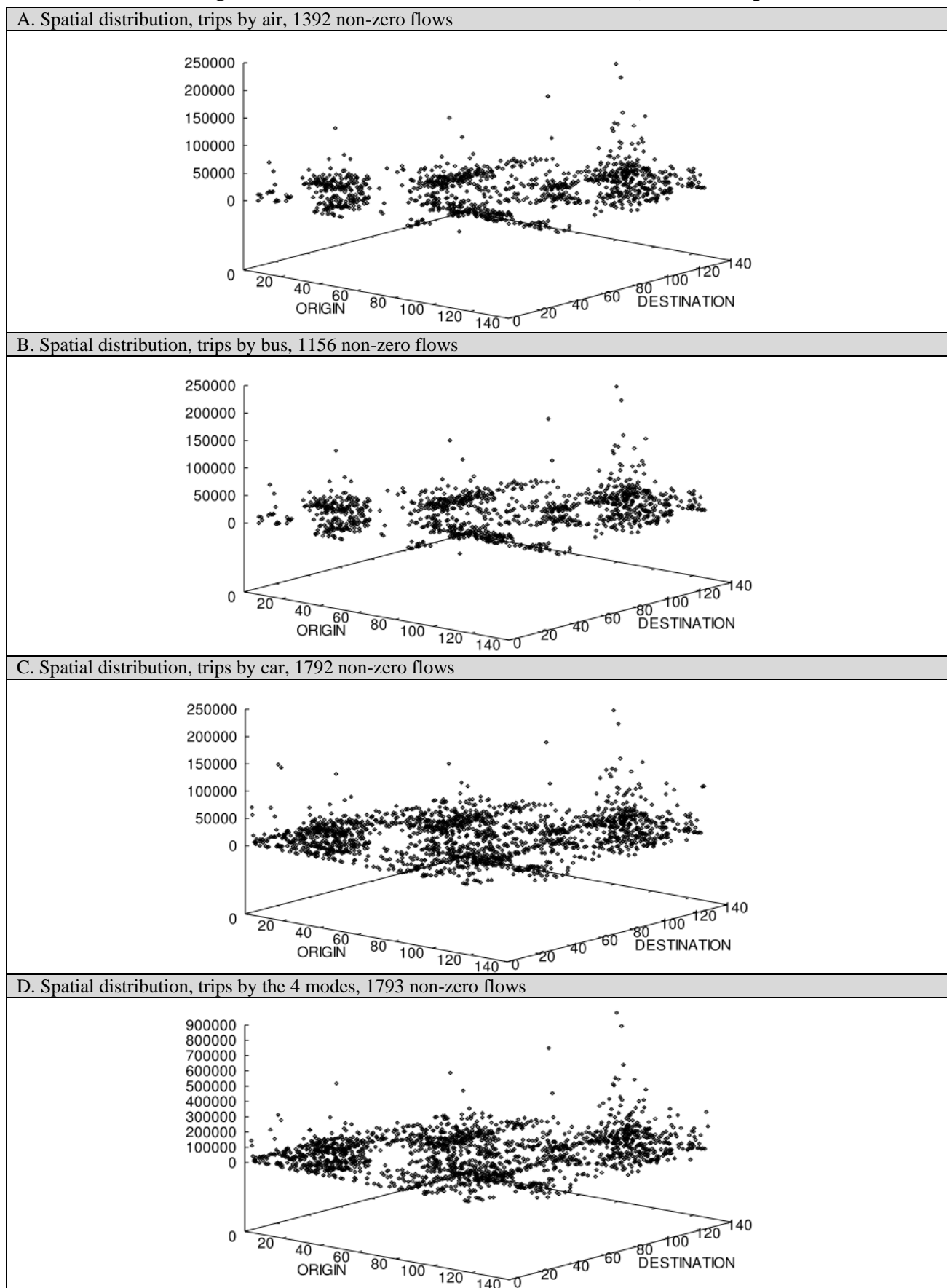
Table 21. QDF demand elasticities for five cores (variable X in RUF$_m$ and in other RUF$_j$)

(44)	**Logit**
η_m	$\beta_U \dfrac{U^{\lambda_U}}{T_{TOT}^{\lambda_{TOT}}} \displaystyle\sum_{j=1}^{M} \beta_{jX} X^{\lambda_{jX}} P_j + \beta_{mX} X^{\lambda_{mX}} \left(1 - P_m\right) - P_j \displaystyle\sum_{j \neq m}^{M} \beta_{jX} X^{\lambda_{jX}}$
(45)	**Standard Dogit**
η_m	$\beta_U \dfrac{U^{\lambda_U}}{T_{TOT}^{\lambda_{TOT}}} \dfrac{(1 + \displaystyle\sum_{j=1}^{M} \theta_j) \displaystyle\sum_{j=1}^{M} \beta_{jX} X^{\lambda_{jX}} \exp(V_j)}{\displaystyle\sum_{j=1}^{M} \left[\exp(V_j) + \theta_j \displaystyle\sum_{j=1}^{M} \exp(V_j) \right]} + \dfrac{A_m(1 - P_m) - P_m \displaystyle\sum_{j \neq m}^{M} A_j}{1 + \theta_m \displaystyle\sum_{j=1}^{M} \exp(V_j - V_m)},$ where $A_m = \left[\beta_{mX} X^{\lambda_{mX}} + \theta_j \displaystyle\sum_{j=1}^{M} \beta_{jX} X^{\lambda_{jX}} \exp(V_j - V_m) \right] \qquad A_j = \left[\beta_{jX} X^{\lambda_{jX}} \exp(V_j - V_m) + \theta_j \displaystyle\sum_{j=1}^{M} \beta_{jX} X^{\lambda_{jX}} \exp(V_j - V_m) \right]$
(46)	**Generalized Dogit**
η_m	$\beta_U \dfrac{U^{\lambda_U}}{T_{TOT}^{\lambda_{TOT}}} \dfrac{\displaystyle\sum_{m=1}^{M} \left[\beta_{mX} X^{\lambda_{mX}} \exp(V_m) + \displaystyle\sum_{j \neq m}^{M} \theta_{mj} \beta_{jX} X^{\lambda_{jX}} \exp(V_j) \right]}{\displaystyle\sum_{m=1}^{M} \left[\exp(V_m) + \displaystyle\sum_{j \neq m}^{M} \theta_{mj} \exp(V_j) \right]} + \dfrac{B_m(1 - P_m) - P_m \displaystyle\sum_{j \neq m}^{M} B_j}{1 + \displaystyle\sum_{j \neq m}^{M} \theta_{mj} \exp(V_j - V_m)},$ where $B_m = \left[\beta_{mX} X^{\lambda_{mX}} + \displaystyle\sum_{j \neq m}^{M} \theta_{mj} \beta_{jX} X^{\lambda_{jX}} \exp(V_j - V_m) \right] \qquad B_j = \left[\beta_{jX} X^{\lambda_{jX}} \exp(V_j - V_m) + \displaystyle\sum_{j \neq m}^{M} \theta_{mj} \beta_{jX} X^{\lambda_{jX}} \exp(V_j - V_m) \right]$
(47)	**Linear Inverse Power Transformation-Logit**
η_m	$\beta_U \dfrac{U^{\lambda_U}}{T_{TOT}^{\lambda_{TOT}}} \left\{ \dfrac{\displaystyle\sum_{m=1}^{M} \left\{ \beta_{mX} X^{\lambda_{mX}} \exp(V_m) \left[\phi_m \exp(V_m) + 1 \right]^{\frac{1}{\phi_m} - 1} \right\}}{\displaystyle\sum_{m=1}^{M} \left\{ \left[\phi_m \exp(V_m) + 1 \right]^{\frac{1}{\phi_m}} - \mu_m \right\}} + \dfrac{C_m(1 - P_m) - P_m \displaystyle\sum_{j \neq m}^{M} C_j}{\left[\phi_m \exp(V_m) + 1 \right]^{1/\phi_m} - \mu_m} \right\},$ where $C_m = \left[\phi_m \exp(V_m) + 1 \right]^{\frac{1}{\phi_m} - 1} \exp(V_m) \beta_{mX} X^{\lambda_{mX}} \qquad C_j = \left[\phi_j \exp(V_j) + 1 \right]^{\frac{1}{\phi_j} - 1} \exp(V_i) \beta_{jX} X^{\lambda_{jX}}$
(48)	**Box-Tukey Inverse Power Transformation-Logit**
η_m	$\beta_U \dfrac{U^{\lambda_U}}{T_{TOT}^{\lambda_{TOT}}} \displaystyle\sum_{m=1}^{M} \beta_{mX} X^{\lambda_{mX}} \exp(V_m) \left[\exp(V_m) + \tilde{\mu}_m \right]^{\tilde{\phi}_m - 1} P_m + D_m(1 - P_m) - P_m \displaystyle\sum_{j \neq m}^{M} D_j,$ where $D_m = \left[\exp(V_m) + \tilde{\mu}_m \right]^{\tilde{\phi}_m - 1} \exp(V_m) \beta_{mX} X^{\lambda_{mX}} \qquad D_j = \left[\exp(V_j) + \tilde{\mu}_j \right]^{\tilde{\phi}_j - 1} \exp(V_j) \beta_{jX} X^{\lambda_{jX}} \exp(E_j)$ and $E_j = \left[\exp(Vj) + \tilde{\mu}_j \right]^{(\tilde{\phi}_j)} - \left[\exp(V_m) + \tilde{\mu}_m \right]^{(\tilde{\phi}_m)}$

---

[135] The detailed derivations are found in Tran & Gaudry (2010a).

11. Appendix C. Coverage of business trips, VIA RAIL 1987 database

Figure 26. Other Corridor 136x136 OD matrices, business trips

A. Spatial distribution, trips by air, 1392 non-zero flows

B. Spatial distribution, trips by bus, 1156 non-zero flows

C. Spatial distribution, trips by car, 1792 non-zero flows

D. Spatial distribution, trips by the 4 modes, 1793 non-zero flows

12. Appendix D. On the coupling term and OD-type consistency with IIA

We have addressed, somewhat marginally, the issue of consistency with IIA arising from the use of only *own LOS* variables in Modal utility functions but have said little about OD-type consistency: as stated in Section 2, our formulation of the QDF sidesteps the issue and assumes that each OD flow is essentially determined by variables that share the indices of the dependent variable. The reason for this is that our emphasis has been on the relevance of form estimation; but we do not want to give the impression that spatial competition is properly modelled in our examples.

Impedance, Distance or Utility in logarithmic form. We have also pointed out in Section 2 that trade and transport models were closely related and have discussed in Section 5.4.A the proper form of the coupling term in QDF transport applications. It is therefore instructive[136] to compare the common part of trade and transport model specifications, namely component (6-A) explaining Total flow by origin-destination pair, fleshed out simply as:

(49) $(\text{Flow}_{ij}) \leftarrow (\text{Activities}_{i \text{ and } j}; \text{Socioeconomic}_{i \text{ and } j}, ; \text{Ease of interaction}_{ij})$

which, since Carey (1858-1859) proposed it in strict Newtonian form to describe spatial human interactions, was in fact first used in transportation by Lill (1891) on Vienna-Brünn-Prague passenger rail flow data without forcing the Distance impedance term to be raised to the square power Newtonian value. It could be argued that the extended multiplicative "Gravity" model applied to human interactions (*e.g.* to trade, transport or communication flows) and usually estimated by Ordinary Least Squares in logarithmic form and without double constraints on origin or destination "trip ends", has since been the most robust model in all of Economics and influential there since trip length and mode choice became matters of formal analysis (*e.g.* Morgenstern, 1936).

Distance was used at first as the typical measure of (dis)-Utility or "impedance" in all applications, but in transport the formal term (3) known to-day as Utility took hold early between 1965 and 1975, as did other refinements such as border effects, all of much later import in trade models. BCT were also used in intercity transport models since 1976, long before their first uses in trade models (Gaudry *et al.*, 1996; Gaudry, 2004). Among early authors, Kau and Sirman (1979), using intercity traffic data for the American State of Georgia, could not reject the multiplicative form of the Distance term, contrary to what most other tests have since found, in transport or, later, in trade.

Form and identification. In trade models, Distance is still almost universally employed as a proxy for transport cost, and in the logarithmic form of (49), sometimes with dire consequences when used as a surrogate for *changing* transport conditions to compare impedance effects from samples pertaining to different moments of time (Buch *et al*., 2004): in those cases, assumed proportionate changes in costs between the samples will be accounted for in the constant but not as desired in the coefficient of the distance term. This problem does not arise if a Box-Cox form is used to transform the distance term: changes over time in distance elasticities will then be proper surrogates of changes in price elasticities. Clearly, gains from moving away from the fixed log-log form go much beyond mere fit —some "distance" effects can only be identified if the log-log form is *not* optimal!

The use of Distance as a proxy for price in time series, say from 1970 until 1995 or 1999 (Ghosh & Yamarik, 2004; Fratianni & Kang, 2006) is therefore incorrect in a gravity trade model of log-log form, even with period dummy variables: Distance then simply reflects basically constant propensities to trade with more or less distant partners, independently from their desired critical role as vehicles of changing transport conditions in explaining increased globalization flows. Therefore Distance has to be used *in addition to* transport costs if log-log trade flow explanations are tested with time-series data, as some authors correctly do (*e.g.* Giuliano *et al*., 2006). And, to the extent

---

[136] This appendix draws heavily from Section 3.2. of Gaudry (2010).

that more sophisticated measures of impedance than distance are in rough proportion to it, any proper comparison of the roles of distance and of its more sophisticated replacements, *even for the same period*, also requires that the form *not be* logarithmic.

But the two issues, that of the measure of impedance and that of its form, should remain distinct from each other and from a third issue, that of the OD-type IIA consistency of (49) where the explanation of a flow from *i* to *j* only involves variables that have *i*, *j*, or *ij* indices. This means that markets, say for tourism, are unrelated: the Demand for going to A is independent of the fare to B.

More than one price: the problem of spatial competition. This last issue has also been addressed in both transportation and trade models, because OD-type consistency with IIA bothers many. A difficulty for all is that there are N^2 flows, or say $(N^2 - N)/2$ if the diagonal is neglected and symmetry imposed. Another is that theory further requires both substitutes and complements: so-called "intervening opportunity" models and their gravity-opportunities generalization (Wills, 1986) were a way of introducing some complementarity among destinations perhaps situated on a given path. We want to discuss a more parsimonious approach that does not require the inclusion of all prices, but only of the relevant ones, and simultaneously allows for complements among flows without opening the door to the collinearity forcefully avoided by diagonal slavery imposed in (49).

The first way to include cross-prices, already mentioned above, of course consists in using trip end constraints on emissions and attractions, a general and well documented practice in transport analysis (*e.g.* Batten & Boyce, 1986). These constraints simultaneously bring transport conditions *all-in* and *force substitution* on the pattern: modifying a link will readjust all others in the same direction to meet the constraints. But taking theory seriously requires Occam's razor here because not all competing OD pairs matter to explain the flow for a particular one, and also requires admitting the possibility of complements, as in classical demand systems. We have argued already that this artificial linkage among flows should be rejected, notwithstanding its success in Distribution models. But other profligate ways developed for trade analysis should also be avoided.

Other *all-in* formulations excluding complementarity among flows. A representative example is provided by Wei (1996) who introduces an accessibility measure defined as a weighted average of distances to all trading partners, in the old fashion of Trip Generation studies devoid of transport conditions as explanatory variables and desperately trying to bring the network back in somehow. Another example is provided by Anderson and Wincoop (2003), based on Anderson (1979), who introduce in the model "multilateral resistance terms in the form of an atheoretic remoteness variable related to distance from all bilateral partners". The latter illumination is an economic version (based on expenditures) of the traffic flow constraints just rejected: instead of imposing an heroic structure on the residuals, it imposes comparably heroic assumptions about the spatial distribution of expenditure shares that never hold in multilateral trade models.

Trade models typically deal simultaneously with final and intermediate goods (not just with final goods) and always exclude service flows and balancing financial flows from the analysis. In the absence of such offsetting items required by balance of payments constraints, it is natural to find, when testing with BCT flow matrices constructed without such accounting constraints: (i) power values of Generation terms (whether based on Population or on Income or Output) systematically different from those of corresponding Attraction terms; (ii) and both of course different from 1. In addition to such practical pound foolishness, these studies simultaneously practice a form of theoretical penny wisdom by imposing substitution between the own flow of interest and all others.

Parsimony without the imposition of substitution. A better way to be systematic is to introduce just the right number of cross utility terms by an analysis of the spatial correlation among residuals. For this, consider explicitly a system where the first two equations reproduce (6-A) and (6-D) and

the third adds a mechanism, equation (52), allowing for spatial autocorrelation of the homoscedastic residual v_t associated with an observation t to be correlated with many values of the same vector v_n, for instance in a case of 2 orders:

$$(50) \qquad y_t^{(\lambda_y)} = \sum_{k=1}^{K} \beta_k \cdot X_{kt}^{(\lambda_{X_k})} + u_t \ .$$

$$(51) \qquad u_t = \left[\exp\left(\sum_m \delta_m Z_{mt}^{(\lambda_{zm})} \right) \right]^{1/2} \bullet v_t$$

$$(52) \qquad v_t = \sum_{l=1}^{2} \rho_l \left(\sum_{n=1}^{N} \widetilde{r}_{l,tn} v_n \right) + w_t$$

where $\widetilde{r}_{l,tn}$ is the typical element of the matrix $\widetilde{R}_l$, a notation which may be clearer in matrix form:

$$(53) \qquad v = \sum_{l=1}^{2} \rho_l \widetilde{R}_l v + w$$

and $\widetilde{R}_l$ is the (row or column) normalized Boolean matrix R. The Boolean matrix expresses an hypothesis concerning the presence of correlation, for instance among first neighbour zones at the origin and destination of the ij flow, or perhaps (based on other assumptions) among some competing or complementary destinations. It is possible to have both substitutes and complements in (52), depending on the nature or source of the potential correlation tested by each $-1 < \rho_l < 1$. One could for instance imagine that different regions might be competing for tourists ($\rho_l > 0$) but that sub-areas within regions could be simultaneously complementary ($\rho_2 < 0$). As only some OD pairs pertaining to non ij flows are selected when spatial autocorrelation is detected in this way, the resulting system is one of diagonal dominance because the cross terms have, with $-1 < \rho_l < 1$, a smaller role than that of diagonal terms in explaining ij flows, as the reader can readily verify by substituting (52) into (51) and the latter into (50).

From diagonal dominance towards a weighted and distributed variant of *all-in*. But this *more or less* all-in structure is also possible with an extension, with the data deciding on the role of distant flows. Assume that one wants to test the influence of neighbours of neighbours, and so on in geometrically declining strength, on the relevant flow: in this case, we redefine the $\widetilde{R}_l$ matrix as:

$$(54) \qquad \widetilde{R}_l = \pi_l \left[I - (1 - \pi_l) \overline{R}_l \right]^{-1} \overline{R}_l, \qquad (0 \prec \pi_l \leq 1) \ ,$$

where the proximity parameter π_l allows endogenisation of the relative importance of near and distant effects considered in the rule of construction of R_l. If π_l equals one, $\widetilde{R}_l$ is equal to $\overline{R}_l$, indicating that only the adjacent neighbours have an impact on the correlations among the associated residuals assumed in defining the matrix of neighbours $\widetilde{R}_l$: this corresponds exactly to a classical case (Ord, 1975). By contrast, as π_l tends towards zero, the near effect is reduced to a minimum in favour of the distant effect. Generally, the π_l weigh the *relative* importance of near and distant effects with a single parameter defining the sharpness or slope of the decline. It is therefore possible, with such a distributed lag, to test for and weigh influences beyond those assumed by the original $\widetilde{R}_l$ structure. This is far more realistic than full *all-in* systems where the balancing constraints treat all OD flows in the same way[137]. Results of such tests are found in Table

---

[137] For details, see Blum *et al.* (1996): this Boolean fit technique is strongly preferable to using functions, for instance of Distance, in the R matrix, a technique that does not yield results invariant to units of measurement of the Distance variable, as pointed out in an urban mode choice example (Bolduc *et al.*, 1989, Footnote on page 369).

22 where one also finds a demonstration of the effect of replacing Distance by the Utility term in Total demand models and of the application of BCT to previously logarithmic specifications.

Table 22. Form, heteroskedasticity and spatial correlation in models of Total passenger transport

Part I		Single-country models for all modes and all trip purposes (domestic flows)					
Gaudry *et al.*, (1994)		Canada 1976 (120 observations)[138]			Germany 1985 (286 observations)[139]		
Column		*1*	*2*	*3*	*4*	*5*	*6*
Case		*Log-log*	*Log-log +1st autoc.*	*4 Box-Cox +1st autoc.*	*Log-log*	*Log-log +1st autoc.*	*4 Box-Cox +1st autoc.*
Variable							
Utility	elasticity	0,63	0,57	0,69	0,40	0,25	0,24
	t-statistic	(42,33)	(25,27)	(24,37)	(9,68)	(5,59)	(5,71)
Other variables		[...]	[...]	[...]	[...]	[...]	[...]
Box-Cox λ		**0,00**	**0,00**	**-0,08**	**0,00**	**0,00**	**0,41**
Spatial autocorrelation ρ		0,00	0,75	0,86	0,00	0,80	0,73
Distributed lag π		--	1,00	0,00	--	0,74	0,49
Log-likelihood		-1318	-1294	-1262	-3057	-3028	-2995
Degrees of freedom Δ		0	2	6	0	2	6

Part II		Joint France (1993-1994) and United Kingdom (1991) model, all modes by trip purpose (domestic flows and flows between them or with 11 other European countries)[140]					
Last (1998)		Vacation trips (6540 observations)			Private trips (8334 observations)		
Column		*7*	*8*	*9*	*10*	*11*	*12*
Variable	*Case*	*Log-log*	*3 Box-Cox*	*3 Box-Cox*	*Log-log*	*2 Box-Cox*	*2 Box-Cox*
Distance	elasticity	-0,23	**-0,20**		-0,72	**-0,57**	
	t-statistic	(-8,71)	(-12,95)		(-22,46)	(-30,00)	
Utility	elasticity			**0,29**			**0,52**
	t-statistic			(12,00)			(26,99)
Other variables		[...]	[...]	[...]	[...]	[...]	[...]
Box-Cox λ		**0,00**	**-0,74**	**0,12**	**0,00**	**-0,57**	**0,31**
Log-likelihood		-65407	-65313	-65390	-80804	-79991	-79995
Degrees of freedom Δ		0	3	3	0	2	2

Part III		Business trips full set of 5711 observations			Business trips subset of 544 observations		
Gaudry *et al.* or Last (1998)							
Column		*13*	*14*	*15*	*16*	*17*	*18*
Case		*Log-log*	*3 Box-Cox +heterosk.*	*3 Box-Cox +heterosk.*	*Log-log*	*8 Box-Cox +heterosk.*	*As (17) +1st autoc.*
Variable							
Distance	elasticity	-0,33	**0,20**				
	t-statistic	(-8,78)	(-11,02)				
Utility	elasticity			**0,22**	0,11	0,15	0,16
	t-statistic			(12,50)	(4,36)	(5,66)	(5,62)
Other variables		[...]	[...]	[...]	[...]	[...]	[...]
Box-Cox λ		**0,00**	**-1,02**	**0,33**	**0,00**	**0,25**	**0,25**
Spatial autocorrelation ρ							0,18
Log-likelihood		-50846	-50762	-50737	-4890	-4858	-4852
Degrees of freedom Δ		0	5	5	0	10	11

---

[138] Results drawn from Table B.2, Columns 1, 3 and 6 of the source paper. The Utility term is constructed from the 4-mode Box-Cox Logit model summarized as Model 8 in Table 7 above. It is specified in RATE format with distinct transformations applied to its four network variables (Price, Speed, Distance, Frequency) but not to its Socio-economic variables.

[139] Results drawn from Table 3, Columns 1, 3 and 5 of the source paper. The utility term is constructed from the 3-mode Box-Cox Logit model summarized as Model 9 in Table 7 above. It is specified in RATE format with distinct transformations applied to its four network variables (Price, Speed, Distance, Frequency) but not to its Socio-economic variables.

[140] Results drawn from Appendices 2, 3 and 4 of the source paper except for the autocorrelation tests which are drawn from Table 11. The Utility term is constructed from the 3-mode Box-Cox Logit models by trip purpose summarized as Models 11-13 in Table 7 above. Each is specified with a single transformation common to Fare and In-vehicle time network variables, except in the Vacation trip model where a BCT is only applied to the Fare term.

The measure and form of the ease of interaction, and the problem of IIA. As this is an appendix, we comment only briefly on the results found in Table 22.

First, greyed cases illustrate the impact of replacing Distance by Utility, defined by (3), while simultaneously allowing for a BCT instead of a logarithmic specification: we note that the effect on elasticities is quite small, no doubt in part because the spatial structure of modal services is closely correlated to Distance and because optimal BCT estimates have relatively little impact on elasticities. This is even clearer when an examination of the six models presented is made: although fit always improves with BCT, the elasticities are never very strongly different from their values in log-log specifications no matter how sophisticated the gravity specification might be. But what of the specific impact of spatial autocorrelation?

How is the Boolean matrix R defined? First neighbours are here defined for any given flow as all those other flows having origins or destinations within 320 km (Columns 2 and 3), or in a range between 100 and 180 km (Columns 5 and 6), or between 100 and 300 km (Column 18). Some results are obtained without tails distributed in accordance with (54) but, in Columns 3 and 6, such significant "long tail" π_l parameters are duly estimated with the algorithm documented in Tran and Gaudry (2008c). Although much better fits are then generally obtained than without these additions, the log-likelihood gains are greatest in models considering all trip purposes together (Columns 5 and 6): they are clearly lower when only the business trip purpose (Column 18) is analyzed. This meets expectations because the substitution among destinations makes less sense for business trips than for trips for other purposes, such as tourism. In all tested cases, flows are substitutes because the estimated signs of the ρ_l say so, not by assumption.

One senses that gains due to the estimation of forms are smaller or less critical than they were for the Modal choice component and that properly accounting for spatial correlation might be more promising than current *all-in* alternatives to introduce substitution, and perhaps eventually complementarity, among OD flows.

13. References

Abraham, C. and P. Coquand (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, 57-76.

Abrantes, P. A. L. and M. R. Wardman (2011). Meta-analysis of UK values of travel time: An update. *Transportation Research A*, 45, 1-17.

Aguirre, I., Cowan, S. and J. Vickers (2010). Monopoly Price Discrimination and Demand Curvature. *American Economic Review* 100, pp. 1601–1615, September.

Aitken, A. C. (1934-1935). On Least Squares and Linear Combination of Observations. *Proceeding of the Royal Society of Edinburgh* 55, pp. 42-48.

Algers, S. (1984). The calibration and application of travel demand models in Sweden. In Florian, M. (ed.). *Transportation Planning Models*, 79-105, North Holland.

Algers, S. (1993). Integrated Structure of Long Distance Travel Behavior Models in Sweden. *Transportation Research Record 1413*, pp. 141-149, Washington D. C.

Algers, S. (2004a). *Utveckling av Sampers 2.1, Del 1 Estimeringen inkl tilläggstester*, Transek AB, Stockholm, 165 p., January.

Algers, S. (2004b). *A note on tests on functional form of Swedish long distance models.* Unpublished manuscript, Department of Traffic and Transport Planning, Royal Institute of Technology, Stockholm, 3 pages, 28th August (revised 18th November 2009).

Algers, S. (2011). Long distance models in Sweden. Provisional summary report, February 2, 8 pages.

Algers, S. and M. Gaudry (1994). *Are Nonlinearity and Segmentation Substitutes, Complements, or Unrelated in Logit Models? The Swedish Intercity Passenger Case*, Publication AJD-28, Agora Jules Dupuit, Université de Montréal, Version 6, 18 p.. www.ajd.umontreal.ca.

Allenby, G. M. and P. E. Rossi (1991). There is no aggregation bias: why macro logit models work. *Journal of Business and Economic Statistics* 9, 1, 1-14.

Anderson, J. E.(1979). A Theoretical Foundation for the Gravity Equation, *American Economic Review* 69, 1, 106-116, 1979.

Anderson, S. P. and A. de Palma (2000). From Local to Global Competition. *European Economic Review* 44, 3, pages 423-448, March.

Anderson, J. E. and E. van Wincoop (2003). Gravity with Gravitas: A Solution to the Border Puzzle, *American Economic Review* 93, 179-192.

Anderson, S. P., Palma, A. de and J.-F. Thisse (1988). A representative consumer theory of the logit model. *International Economic Review* 29, 3, 461-466, August.

Andersson, J. and J. Ubøe (2010). *Some aspects of random utility, extreme value theory and multinomial logit models.* Discussion Paper FOR 1 3010, Department of finance and management science, Norwegian School of Economics and Business Administration, Bergen, January.

Andrikopoulos, A. A. and J. A. Brox (1990). Canadian intercity passenger transportation: a simultaneous equations approach. *International Journal of Transport Economics* 17, 3, 311-328, October.

Anscombe, F. J. (1961). Examination of residuals. *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, 1-37.

Anscombe, F. J. and J. W. Tukey (1954). *The criticism of transformations*. Paper presented before the Americal Statistical Association and Biometric Society, Montreal.

Arduin, J.-P. (1989). *Modèles de prévision de trafic utilisés par SOFRERAIL*. SOFRERAIL, Paris, 17 pages.

Arduin, J.-P. (1993). *Évaluation des grands projets ferroviaires: les modèles de prévision de trafic.* Société Nationale des Chemins de Fer Français (SNCF), 19 p., Presented at Sixièmes Entretien du Centre Jacques-Cartier, Colloque Villes et T.G.V., Lyon, 5-11 December.

Armington, P. (1969). A Theory of Demand for Products Distinguished by Place of Production. *International Monetary Fund Staff Papers*, 159-176.

Atkinson, A. C. (1983). Diagnostic regression analysis and shifted power transformations. *Technometrics* 25, 1, 23-33, February.

Axhausen, K. W., Hess, S., König, A., Abay, G., Bates, J. J. and M. Bierlaire (2008). Income and distance elasticities of values of travel time savings: new Swiss results. *Transport Policy* 15, 3, pages 173–185.

Azzalini, A. (1994). Logistic regression for autocorrelated data with application to repeated measures. *Biometrika* 81, 767–75.

Batten, D. F. and D. E. Boyce (1986). Spatial Interaction, Transportation, and Interregional Commodity Flow Models. Ch. 9 in Nijkamp, P. (ed.) *Handbook of Regional and Urban Economics, Vol. 1 Regional Economics*, pp. 357-406, North Holland.

Ben-Akiva, M. (1974). *Note on the specification of a logit model with utility functions that include attributes of competing alternatives*. Working Paper, Department of Civil Engineering, MIT, Cambridge, MA, 10 pages, December.

Ben-Akiva, M. (1977). *Choice models with simple probabilistic choice set generation processes*. Department of Civil Engineering, M.I.T., November.

Ben-Akiva, M. and J. Swait (1984). *Choice models with simple probabilistic choice set generation processes.* Submitted for publication to Transportation Research B, 21 p. May.

Ben-Akiva, M. and J. Swait (1987). Incorporating random constraints in discrete models of choice set generation. *Transportation Research B* 21, 2, 91-112.

Ben-Elia, E. and Y. Shiftan (2010). Which road do I take? A learning-based model of route-choice behavior with real-time information. *Transportation Research Part A*. 44, 4, 249-264, May.

Berkson, J. (1944). Application of the logistic function to bio-assay. *Journal of the American Statistical Association* 39, 357-365.

Berkson, J. (1950). Are There Two Regressions? *Journal of the American Statistical Association* 45, 164-180.

Bernouilli, D. (1738). Specimen theoriae novae de mensura sortis. *Commentarii academiae scietiarum imperialis Petropolitanae*, V, St. Petersburg, 175-192.

Bernouilli, D. (1954). Exposition of a new theory of the measurement of risk. *Econometrica* 22, 23-36.

Berry, S., Levinsohn, J. and A. Pakes (2004). Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market. *Journal of Political Economy* 112, 1, 68-105.

Beser, M. and S. Algers (2001). SAMPERS — The new Swedish national travel demand forecasting tool. In L. Lundqvist and L.-G. Mattsson (eds). *National Transport Models: Recent Developments and Prospects,* pp. 101–118. Advances in Spatial Science, Springer.

Beuthe, M. (2010). Users' benefit valuation and the specification of demand functions in situation of congestion. GTM – FUCAM, Louvain School of Management, Catholic University of Mons, 23 p.

Beuthe M. (2010), Users' benefit valuation and the specification of demand functions in a situation of congestion. Ch. 8, in Van de Voorde, E. and Th. Vanelslander (eds), *Applied Transport Economics - A Management and Policy Perspective,* 399-423, De Boeck, Brussels, October.

Bhat, C. (1995). A Heteroscedastic Extreme Value Model of Intercity Travel Mode Choice. *Transportation Research B* 29, 6, 471-483.

Bierlaire, M. (2003). BIOGEME: a free package for the estimation of discrete choice models. *Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Switzerland.

Bierlaire, M. (2008). *An introduction to BIOGEME Version 1.6.* www.biogeme.epfl.ch.

Bierlaire, M., Lotan, T. and P. Toint (1995). *On the Overspecification of Multinomial and Nested Logit Models Due to Alternative Specific Constants*. Rapport 95/12, Publications du département de mathématiques, Facultés Universitaires de Namur, 14 p., October 24.

Blardone, D. (2007). Les préférences des chargeurs pour leurs transports à travers les Pyrénées. *Notes de synthèse du SESP* N[0] 164, 11 p., Janvier-Février-Mars.

Blayac, T. and A. Causse (2001). Value of travel time: a theoretical legitimization of some nonlinear representative in discrete choice models, *Transportation Research B*, 35, pp. 391-400.

Blum, U. C., Bolduc, D. and M. Gaudry (1996). From Correlation to Distributed Contiguities: a Family of AR-C-D Autocorrelation Processes. *Zeitschrift für Regionalwissenschaft* 16/17, 99-128.

Bolduc, D. (1999). A practical technique to estimate multinomial probit models in transportation. *Transportation Research B, 33, 1*, pages 63-79, February.

Bolduc, D., Dagenais, M. G. and M. Gaudry (1989). Spatially Autocorrelated Errors in Origin-Destination Models: A New Specification Applied to Aggregate Mode Choice. *Transportation Research B* 23, 5, 361-372.

Bordley, R. F. (1989). Relaxing the Loyalty Condition in the Colombo/Morrison Model. *Marketing Science*, 8, 1, 100-103.

Bordley, R. F. (1990). The Dogit Model is Applicable even without Perfectly Captive Buyers. *Transportation Research B*, 24, 4, 315-323.

Borenstein, S, and N. L. Rose (1994). Competition and Price Dispersion in the U.S. Airline Industry. *Journal of Political Economy* 102. 4, 653-683.

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

Box, G. E. P. and P. W. Tidwell (1962). Transformation of the Independent Variables. *Technometrics* 4, 531-550.

Brown, H. (1980). Estimating the Dogit Model Using Aggregate Data. Department of civil engineering, University of Melbourne, 7 pages, June.

Bruzelius, N. (1979). *The Value of Travel Time*. Croom Helm, London, 222 pages.

Buch, C. M., Kleinert, J. and F. Toubal (2004). The distance puzzle: on the interpretation of the distance coefficient in gravity equations, *Economics Letters* 83, 293-298.

Carey, H. C. (1858-1859). *Principles of Social Science*, Lippincott, Philadelphia.

Cascetta, E., Nuzzolo, A., Russo, F. and A. Vitetta (1996). A New Route Choice Logit Model overcoming IIA Problems: Specification and Some Calibration Results for Interurban Networks. In J. B. Lesort (ed.) *Transportation and Traffic Theory: Proceedings of the 13th International Symposium on Transportation and Trafic Theory*, Lyon, France, 15 p.

Chateauneuf, A. (1999). Comonotonicity axioms and RDEU theory for arbitrary consequences. *Journal of Mathematical Economics*, 32, 21-45.

Chateauneuf, A. and M. Cohen (1994). Risk seeking with diminishing marginal utility in a non-expected utility model, *Journal of Risk and Uncertainty* 9, 77-91.

Christensen, L. R., Jorgenson, D. W. and L. J. Lau (1971). Conjugate Duality and the Transcendental Logarithmic Function. *Econometrica* 39, 4, 255-256, July.

Chew, S., Karni, E. and Z. Safra (1987). Risk aversion in the theory of expected utility with rank dependent preferences. *Journal of Economic Theory* 42, 370-381.

Cho, J. S., Ishida, I. (2012). Testing for the effects of omitted power transformations. Economics Letters 117, 1, 287-290, October.

Chu, Y.-L. (1990). Combined trip distribution and trip assignment model incorporation captive travel behaviour. *Transportation Research Record* 1285, 70-77.

Chu, Y.-L. (2011). Distribution and assignment of both compulsory and discretionary traffic. Paper 11-3694, TRB Annual Meeting, Washington.

Cirillo, C. (2005). *Optimal Design for Mixed Logit Models*. Groupe de recherche sur les transports, Facultés Universitaires Notre-Dame de la Paix, Namur, 11 p., Association for European Transport and contributors.

Cléroux, R., Florian, M. and S. Galarneau (1981). *A modal choice function for Winnipeg – calibration and validation*. Publication CRT-135, Centre de recherche sur les transports, Université de Montréal.

Cohen, M. and J. M. Tallon (2000). Décisions dans le risque et l'incertain: l'apport des modèles non additifs, *Revue d'Economie Politique*, 110(5), 631-681.

Cox, D. R. (1970). *Analysis of binary data*. Methuen, London.

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.

Cramer, J. S. (2003). *The origins and development of the logit model*. University of Amsterdam and Tinbergen Institute, Amsterdam, 19 p.

Dagenais, M. G. (1986). *A generalized share model with binding constraints.* Unpublished manuscript, Département de sciences économiques, Université de Montréal, and Groupe de Recherche en Économie Quantitative et Économétrie (GREQE), Marseille, 18 pages, November. Presented at the First Annual Meeting of the European Economic Association, Vienna, August 19-31, 1986. Forthcoming as Publication AJD-137, Agora Jules Dupuit, Université de Montréal, www.e-ajd.net .

Dagenais, M. G. and J.-M. Dufour (1994). Pitfalls of Rescaling Regression Models with Box-Cox Transformations. *The Review of Economics and Statistics* 76, 3, 571-75, August.

Dagenais, M. G. and M. Gaudry (1986). Can Aggregate Direct Travel Demand Models Work? *Proceedings of the World Conference on Transport Research*. The Centre for Transportation Studies, University of British Columbia, Vancouver, 1669-1676.

Dagenais, M. G., Gaudry, M. and T. C. Liem (1987). Urban Travel Demand: the Impact of Box-Cox Transformations with Nonspherical Residual Errors. *Transportation Research B* 21, 6, 443-477.

Dagenais, M. G., Gaudry, M. and C. L. Tran (1982a). *Testing the Generalized Dogit mode choice model: a progress report.* Unpublished manuscript, Département de sciences économiques, Université de Montréal, 14 p. Presented at the International Atlantic Economic Conference, February.

Dagenais, M. G., Gaudry, M. J. I. and T.C. Liem (1982b), Testing the Generalized Dogit Mode Choice Model. Abstract. *Atlantic Economic Journal,* 60, September.

Daly, A. (2008). *The relationship of cost sensitivity and trip length.* RAND Europe and Institute for Transport Studies, University of Leeds, 21 p. September.

Daly, A. (2010). *Cost Damping in Travel Demand Models*. Department for Transport Report TR-717-DFT, 42 pages, 13th January. www.dft.gov.uk/pgr/economics/rdg/costdamping.

Davidson R. and J. G. MacKinnon (1984). Convenient specification tests for logit and probit models. *Journal of Econometrics* 25, 241-262.

Davidson R. and J. G. MacKinnon (1993). Transforming the Dependent Variable. Ch. 14 in *Estimation and Inference in Econometrics*. Oxford University Press.

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

Dionne, G., Gagné, R. and C. Vanasse (1992). *A statistical analysis of airline accidents in Canada 1976-1987.* Publication CRT-811, Centre de recherche sur les transports, Université de Montréal, 57 p..

Dionne, G., Gagné, R., Gagnon, F. and C. Vanasse (1997). Debt, moral hazard and airline safety: An empirical evidence. *Journal of Econometrics* 79, 2, 379-402.

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

Domencich, T. A. and G. Kraft (1970). *Free Transit*. Heath Lexington Books, Lexington, Massachusetts.

Donnea, F. X. de (1971). *The determinants of transport mode choice in Dutch cities.* Rotterdam University Press, 230 pages.

Djurisic, M. (2000). *Modélisation des prix de quatre types de vraquiers d'occasion: une approche structurelle autorégressive*. 28 p., avril. Publication AJD-5, Agora Jules Dupuit, Université de Montréal, www.e-ajd-net.

Draper, N. R. and D. R. Cox (1969). On Distributions and Their Transformation to Normality. *Journal of the Royal Statistical Society B* 31, 3, 472-476.

Dufour, J.-M. and E. Renault (1998). Short run and long run causality in time series. *Econometrica* 66, 5, 1099-1125.

Dufour, J.-M. and A. Taamouti (2010). Short and long run causality measures: Theory and inference source. *Journal of Econometrics*. 154, 1, 42-58, January.

Durbin, J. and G. S. Watson (1950). Testing for Serial correlation in Least Squares Regression. *Biometrica* 37, 409-428.

Durbin, J. and G. S. Watson (1951). Testing for Serial correlation in Least Squares Regression. *Biometrica* 38, 159-178.

ECORYS Transport, Macharis, C., Haezendonck, E., Veldman, S., Bückmann, E. and M. Flier (2004). *Ontwikkeling Marktaandeelmodel Containersector*. ECORYS Transport, 134 p. www.proses.nl/scripts/rapportUitwerking.php?rID=67

Ekbote, D. and R. Laferrière (1994). Demand Model Developments to Assess High Speed Passenger Train Markets on Windsor to Quebec City Corridor. *Proceedings of the 20th Annual Meeting, Canadian Transportation Research Forum/Le Groupe de recherches sur les transports au Canada*, 598-609. Publication AJD-55, Agora Jules Dupuit, Université de Montréal. www.ajd.umontreal.ca.

Ellis, R. H. and P. R. Rassam (1970). *National Intercity Travel: Development and Implementation of a Demand Forecasting Framework*. U.S. Dept. of Transportation Contract T-8 542, Modification No. 1, National Technical Information Service PB 192 455, U. S. Dept. of Commerce, March.

Erlander, S. and N. F. Stewart (1990). *The gravity model in transportation analysis − Theory and extensions*. VSP, Utrecht.

Fratianni, M. and H. Kang (2006). Heterogeneous distance–elasticities in trade gravity models, *Economics Letters* 90, 68-71.

Fridstrøm, L. (1992). Causality – Is it all in your mind? An inquiry into the definition and measurement of causal relations. In Ljones, O., Moen, B. and L. Østby (eds): *Mennesker og modeller – livsløp og kryssløp. Social and Economic Studies* 78, Central Bureau of Statistics, Oslo, pages 103-133.

Fridstrøm, L. (2000a). The TRULS-1 model for Norway. Ch. 4 in Gaudry, M. and S. Lassarre, (dir.), *Structural Road Accident Models: The International DRAG Family*, pp. 97-126, Elsevier Science, Oxford.

Fridstrøm, L. (2000b). The IRPOSKML procedure of estimation. Ch. 13 in Gaudry, M. and S. Lassarre, (dir.), *Structural Road Accident Models: The International DRAG Family*, pp. 325-334, Elsevier Science, Oxford.

Fridstrøm, L. and A. Madslien (2002). A Stated Preference Analysis of Wholesalers' Freight Choice. Ch. 9 in Danielis, R. (ed.), *Demanda di Trasporto Merci e Preferenze Dichiarate – Freight Transport Demand and Stated Preference Experiments*. Franco Angeli, Milano, pp. 223-250.

Fry, T. R. L. (1988). *On the Use of some Burr Family Distributions in Econometrics*. Ph.D. thesis, Faculty of Economic and Social Studies, University of Manchester.

Fry, T. R. L. and M. N. Harris (1995). *The Size and Power Properties of Combining Choice Set Partition Tests for the IIA Property in the Logit Model.* Department of Economics, Monash University.

Gallant, A. R. (1981). On the Bias in Flexible Forms and an Essentially Unbiased Form: the Fourier Flexible Form. *Journal of Econometrics* 15, 211-245.

Gaudry, M. (1980). Dogit and Logit Models of Travel Mode Choice in Montreal. *Canadian Journal of Economics/Revue Canadienne d'Économique* 13, 2, 268-279.

Gaudry, M. (1978). *The Inverse Power Transformation Logit and Dogit Mode Choice Models*. Publication CRT-96, Centre de recherche sur les transports, Université de Montréal, 15 p., April.

Gaudry, M. (1981). The Inverse Power Transformation Logit and Dogit Mode Choice Models. *Transportation Research B* 15, 2, 97-103.

Gaudry, M. (1985). Modèles agrégés et désagrégés à forme variable: résultats sur Montréal et Paris. *Transports* 304, 288-293.

Gaudry, M. (1990). *Three Families of Mode Choice Models Applicable to Intercity Travel Demand*. Discussion Paper N° 13, John Deutsch Institute for the Study of Economic Policy, Queen's University, Kingston, 26 p., July.

Gaudry, M. (1993). Asymmetric Shape and Variable Tail Thickness in Multinomial Probabilistic Responses to Significant Transport Service Level Changes. In Comtois, C. and F. Bingquan (eds.). *Proceedings of the '92 International Conference on Urban Land Use and Transport System ICULTS '92*. Publication CRT-910, Centre de recherche sur les transports, Université de Montréal, April.

Gaudry, M. (1994). *Tópicos avanzados en analisis de demanda de transporte: rol de la forma funcional.* Paper presented at the Seminar organized by the Secretaría Ejecutiva de la Comisión de Planificación de Inversiones en Infrastructura de Transporte (SECTRA), 16 pages, Santiago de Chile, 7th January.

Gaudry, M. (1998). Key Substitution-Complementarity Features of Travel Demand Models, with Reference to Studies of High Speed Rail Interactions with Air Services. In COST 318, *Interactions between High-Speed Rail and Air Passenger Transport*, Final Report of the Action, European Commission, Directorate General Transport, Brussels, 204-238.

Gaudry, M. (2001). Tests of Nonlinearity, Modal Captivity and Spatial Competition within the STEMM Multicountry Application for Passengers. Ch. 14 in Lundqvist, L. and Mattsson, L-G. (eds), *National Transport Models*, 170-187, Springer Verlag.

Gaudry, M. (2004). The Robust Gravity Form in Transport and Trade Models. *Zeitschrift fűr Kanada-Studien* 24, 1, Band 44, 144-161.

Gaudry, M. (2005). *The Value of Box-Cox Transformations in the Standard Box-Cox Logit*. Unpublished note, Version 1, 8 pages, June 16.

Gaudry, M. (2010). Improved Modelling of Competition among Airports through Flexible Form and Non Diagonal Demand Structures Explaining Flows registered within a New Traffic Accounting Matrix. Ch. 13 in Forsyth, P., Gillen, D., Müller, J., and H.-M. Niemeier (eds), *Competition in European Airports*, p. 197-236, Ashgate. http://www.ashgate.com/isbn/9780754677468 . Publication AJD-121, Agora Jules Dupuit, Université de Montréal. www.ajd.umontreal.ca.

Gaudry, M., Blum, U. and J. McCallum, (1996). A First Gross Measure of Unexploited Single Market Integration Potential. S. Urban (ed.), *Europe's Challenges*, 449-461, Gabler Verlag.

Gaudry, M. et al. (2001). *TRIO, Version 2.0*. Agora Jules Dupuit, Université de Montréal. www.e-ajd.net.

Gaudry, M. and M. G. Dagenais (1977). *The Dogit Model*. Publication CRT-82, Centre de recherche sur les transports, Université de Montréal. 14 pages. October.

Gaudry, M. and M. G. Dagenais (1981). *The Generalized Dogit Mode Choice Model*. Publication CRT-107, Centre de recherche sur les transports, et Cahier #7826, Département de sciences économiques, Université de Montréal, 20 p.

Gaudry, M. and M. G. Dagenais (1979a). The Dogit Model. *Transportation Research B* 13, 2, 105-112.

Gaudry, M. and M. G. Dagenais (1979b). Heteroscedasticity and the Use of Box-Cox Transformations. *Economics Letters* 2, 3, 225-229.

Gaudry, M. and R. Laferrière (1989). The Box-Cox Transformation: Power Invariance and a New Interpretation. *Economics Letters* 30, 27-29.

Gaudry, M. and A. Le Leyzour (1994). *Improving a Fragile Linear Logit Model Specified for High Speed Rail Demand Analysis in the Quebec-Windsor Corridor of Canada*. Publication CRT-988, Centre de recherche sur les transports ; Cahier 9413, Département de sciences économiques ; Publication IEA-94-07, École des Hautes Études Commerciales, Université de Montréal. Publication AJD-120. www.ajd.umontreal.ca.

Gaudry, M. and E. Quinet (2010). *Track Wear and Tear Cost by Traffic Class: Functional Form, Zero Output Levels and Marginal Cost Pricing Coverage on the French Rail Network*, Publication AJD-30, Agora Jules Dupuit, Université de Montréal, 50 p., April. www.ajd.umontreal.ca.

Gaudry, M. J. I. and M. J. Wills (1978). Estimating the Functional Form of Travel Demand Models. *Transportation Research* 12, 4, 257-289, 1978.

Gaudry, M. J. I. and M. J. Wills (1979). Testing the Dogit Model with Aggregate Time Series and Cross Sectional Travel Data. *Transportation Research B* 13, 2, 155-166.

Gaudry, M., Blum, U. and T. Liem (2000). Turning Box-Cox including Quadratic Forms in Regression. Ch. 14 in Gaudry, M. and S. Lassarre, (eds). *Structural Road Accident Models: The International DRAG Family*. Pergamon, Elsevier Science, Oxford, 335-346.

Gaudry, M., Briand, A, Paulmyer, I. and C.-L. Tran (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.

Gaudry, M., Heinitz, F., Last, J. and B. Mandel (1998). *Methodological Developments within the Quasi-Direct Format Demand Structure: the Multicountry Application for Passengers MAP-1*. Publication CRT-98-46, Centre de recherche sur les transports, Université de Montréal, Working Paper N° 9815, Bureau d'économie théorique et appliquée (BETA), Université Louis Pasteur, 67 p., octobre. www.e-ajd.net.

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

Gaudry, M., Mandel, B. and W. Rothengatter (1994). *Introducing Spatial Competition through an Autoregressive Contiguous Distributed (AR-C-D) Process in Intercity Generation-Distribution Models for Canada and Germany*. Publication AJD-14, Agora Jules Dupuit, Université de Montréal. www.e-ajd.net.

Ghosh, S. and S. Yamarik (2004). Does trade creation measure up? A reexamination of the effects of regional trading arrangements, *Economics Letters* 82, 213-219.

Gillen, D. W. and D. J. Cox (1979). Assumed versus estimated functional form in disaggregate mode choice models. *Regional Science and Urban Economics* 9, 185-195.

Gillen, D., Harris, R. and T. Oum (1999). Evaluating Air Liberalization Agreements: An Integration of Demand Analysis and Trade Theory. Ch. 15 in Gaudry, M. and R. Mayes (eds). *Taking Stock of Air Liberalization*, Kluwer Academic Press, p. 229-250.

Gillen, D., Hinsch, H., Mandel, B. and B. Wolf. (2001). *The Impact of Liberalizing International Aviation Bilaterals on the Northern German Region*. 246 p., Ashgate.

Giuliano, P., Spilimbergo, A., and G. Tonon (2006). *Genetic, Cultural and Geographical Distances*. Discussion Paper No. 2229, Institute for the Study of Labor (IZA), Bonn, 37 p., July.

Givon, M. and D. Horsky (1978). Market share models as approximators of aggregated heterogeneous brand choice behavior. *Management Scien*ce 24, 13, 1404-1416.

Goldfeld, S. M. and R. E. Quandt (1965). Some tests for homoskedasticity. *Journal of the American Statistical Association,* 60, 539-547.

Grandmont, J.-M. (1992). Transformations of the commodity space, behavioural heterogeneity and the aggregation problem. *Journal of Economic Theory* 57, 1-35.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–459.

Granger, C. W. J. (2008). Non-Linear Models: Where Do We Go Next - Time Varying Parameter Models? *Studies in Nonlinear Dynamics & Econometrics* 12, 3, Article 1. http://www.bepress.com/snde/vol12/iss3/art1.

Granger, C. W. J. (2010). Some thoughts on the development of cointegration. *Journal of Econometrics* 158, 3-6.

Granger, C. W. J. and T-H Lee (1999). The effect of aggregation on nonlinearity. *Econometric Reviews* 18, 3, 259-269.

Gronau, R. (1970). *The Value of Time in Passenger Transportation: the Demand for Air Travel*. Occasional Paper 109, National Bureau of Economic Research, 74 pages.

Heckman, J. and S. Polachek (1974). Empirical evidence on the functional form of the earnings-schooling relationship. *Journal of the American Statistical Association* 69, 346, 350-354. June.

Hensher, D. (1976). Market segmentation as a mechanism in allowing for variability of traveller behaviour. *Transportation* 5, 257-284.

Hensher, D. A. (1982). On the equivalence of modified logit models compressed/saturated logit and binary dogit. *Transportation* 11: 87-90.

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

Hensher, D. A. and L. W. Johnson (1979b). External structure of variables in individual choice models of travel demand. *International Journal of Transport Economics* 6, 51-62.

Hensher, D. A. and L. W. Johnson (1981). Behavioural response and form of the representative component of the indirect utility function in travel choice models. *Regional Science and Urban Economics* 11, 559-572.

Hess, S., Erath, A. and K. Axhausen (2008). Estimates of the valuation of travel time savings in Switzerland obtained from pooled data. *Transportation Research Record* 2082, pages 43–55.

Hicks, J. R. (1939). *Value and Capital*. Second edition, Clarendon Press, Oxford.

Hildenbrand, W. (1983). On the Law of Demand. *Econometrica* 51, 997-1019.

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

Hivert, L., Orfeuil, J.-P. and P. Troulay (1988). *Modèles désagrégés de choix modal: réflexions méthodologiques autour d'une prévision de trafic*. Rapport INRETS n[0] 67, Institut National de Recherche sur les Transports et leur Sécurité, Arcueil, 72 pages, juin.

Hjorth, K, and M. Fosgereau (2011). Using prospect theory to investigate the low value of travel time for small time changes. Paper presented at the Kuhmo Nectar Conference, Stockholm, 24 pages, June.

Jackman, S. (1998). *Time Series Models for Discrete Data: solutions to a problem with quantitative studies of international conflict*. Department of Political Science, Stanford University, 39 pages, July 2.

Jara-Diaz, S. (2007). *Transport Economic Theory*. Emerald.

JLR Conseil and Stratec (2005). *Enquête de préférences déclarées sur le choix modal en transport de marchandises transpyrénéen.* Rapport final, Ministère de l'équipement D.T.T./S.l.1, 73 p., 31 août 2005.

Johnson, L. W. (1977). Stochastic parameter estimation: an annotated bibliography. *International Statistical Review* 45, 257-272.

Johnson, L. W. (1978). Regression with Random Coefficients. *Omega* 6, 71-81.

Johnson, L. W. (1979). An introduction to generalised functional form and random coefficients in transport modelling. *Environment and Planning A* 11, 1929-1037.

Johnson, N. L. and S. Kotz (1970). *Continuous Univariate Distributions-1*. John Wiley and Sons.

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

Kahneman, D. and A. Tversky (1979). Prospect theory: an analysis of decision under risk. *Econometrica,* 47, 263-291.

Kau, J. B. and C. F. Sirmans (1979). The Functional Form of the Gravity Model, *International Regional Science Review* 4, 2, 127-136.

Kessel, P., Röling, W. and H.-P. Kienzler (1986). *Untersuchungsbericht zum Forschungsprojekt der MVP Versuchs- und Planungsgesellschaft für Magnetbahnsysteme, Modellentwicklung zur Prognose der Verkehrsnachfrage auf Magnetbahnsystemen*. Beratergruppe für Verkher+Umwelt (BVU), Freiburg im Bresgau.

Kidokoro, Y. (2007). *A Consistent Representative Consumer Framework for Discrete Choice models with Endogenous Total Demand*. Paper presented at 10e Journée Transports, Paris.

Koppelman, F. S. (1981). Non-linear utility functions in models of travel choice behavior. *Transportation* 10, 127-146.

Korenman, S. D., Ouellette, P. and J. Wooldridge (1987). *A new test of aggregate demand curvature properties*. Cahier 0188, Centre de recherche et développement en économique, Université de Montréal, 1988.

KPGM Peat-Marwick and F. S. Koppelman (1990). *Proposal for Analysis of the Market Demand for High Speed Rail in the Québec/Ontario Corridor*. Report presented to the Québec/Windsor Corridor High Speed Rail Task Force, KPGM.

Kraft, G. (1963). *Demand for Intercity Passenger Travel in the Washington-Boston Corridor*. Part V, Northeast Corridor Project, U.S. Department of Transportation.

Kulash, D. J., Kemp, M. A. and R. G. McGillivray (1972). *Aggregation and disaggregate behavioural models*. Working Paper 708-74, The Urban Institute, 17 pages. December.

Laferrière, R. (1988). *Une Agrégation Nouvelle des Itinéraires de Transport Aérien (ANITA).* Publication CRT-574, Centre de recherche sur les transports, Université de Montréal, 106 p., février.

Laferrière, R. (1993a). *HSR in the Quebec-Windsor Corridor: a threshhold analysis with linear and nonlinear RSG models*. Presented to SC Stormont Corporation, 138 p., March 23.

Laferrière, R. (1993b). *Les effets de seuil un train à haute vitesse dans le Corridor Québec-Windsor*. Publication AJD-20, Agora Jules Dupuit, Université de Montréal. Paper presented at the Sixièmes Entretiens du Centre Jacques Cartier, Lyon, December, 14 pages. www.e-ajd.net.

Laferrière, R. (1994). A Travel Demand Forecasting Dilemma: To Pivot or Not? *The Logistics and Transportation Review* 30, 1, 21-29. www.e-ajd.net.

Laferrière, R. (1999). The Air Network Itinerary and Trip Aggregation (ANITA) Model. Ch. 14 in Gaudry, M. and R. R. Mayes (eds), *Taking Stock of Air Liberalization*, Kluwer Academic Press, pages 223-228.

Laferrière, R. and M. Gaudry (1992). Testing the Linear Inverse Power Transformation Logit Mode Choice Model. *Demand, Traffic and Network Modelling*, Selected Proceedings of the Sixth World Conference on Transport Research, Vol II, 739-750.

Lapparent, M. de (2002). *Nonlinearities in the Valuation of Travel Times at the Individual Level*. Publication AJD-68, Agora Jules Dupuit, 43 pages, Université de Montréal. www.e-ajd.net.

Lapparent, M. de (2003). *Individual Demand for Travel Modes and Valuation of Time Attributes within the Regular Journey-to-Work Framework.* Publication AJD-70, Agora Jules Dupuit, Université de Montréal, 28 p, January. www.e-ajd.net.

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* 14, 3, 166-178. 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).

Lapparent, M. de, Frei, A. and K. W. Axhausen (2009). *Choice of mode for long distance travel: current SP-based models from three European countries*. 38 pages, July. Paper presented at the European Transport Conference, Leeuwenhorst, October.

Lapparent, M. de, Palma, A. de and C. Fontan (2002). *Nonlinearities in the Valuation of Travel Attributes.* Publication AJD-69, Agora Jules Dupuit, Université de Montréal, 22 p., October. www.e-ajd.net.

Lave, L. L. (1972). The Demand for Intercity Passenger Transportation. *Journal of Regional Science* 12, 1, 71-84.

Leamer, E. E. (1978a). Regression selection strategies and revealed priors. *Journal of the American Statistical Association* 73, 580–587.

Leamer, E. E. (1978b). *Specification searches: ad hoc inference with nonexperimental data*. John Wiley, New York.

Leonardi, G. (1982). *The Structure of Random Utility Models in the Light of the Asymptotic Theory of Extremes*. WP-82-91, International Institute for Applied Systems Analysis, 45 p., September.

Leonardi, G. (1984). The Structure of Random Utility Models in the Light of the Asymptotic Theory of Extremes. In M. Florian (ed). *Transportation Planning Models*, North-Holland, 107-133.

Lerman, S. R. and J. J. Louvière (1978). On the use of functional measurement to identify the functional form of the utility expression in travel demand models. *Transportation Research Record* 673, 78-86.

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin *et al.* (eds). *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, Stanford University Press, pp. 278-292.

Levin, I., Louvière, J., Meyer, R. and D. Hanely (1980). Individual decision processes in mode choice, *Transportation Research Record*. Paper presented at the Annual Meeting of the Transportation Research Board, January.

Lill, E. (1891). *Das Reisegesetz und zeine Anwendung auf den Eisenbahnverkehr*, Vienna.

Lisco, T. E. (1967). *The Value of Commuter's Travel Time: A Study in Urban Transportation*. Ph.D. Dissertation, Department of Economics, The University of Chicago, Illinois, June.

Luce, R. D. (1959). *Individual Choice Behavior. A theoretical Analysis*. New York, Wiley.

Ley, E. (2006). Statistical inference as a bargaining game. *Economics Letters* 93, 142-149.

Mackie, P., Wardman, M., Fowkes, A., Whelan, G., Nellthorp, J. and J. Bates (2003). *Values of travel time savings in the UK*. Technical report, Institute of Transport Studies, University of Leeds, Leeds, UK.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. Econometric Society Monographs in Quantitative Economics, Cambridge University Press.

Mandel, B. (1992). *Schnellverkehr und Modal Split*. 220 p., Nomon Verlaggesellschaft, Baden-Baden.

Mandel, B. (1999). The interdependency of airport choice and travel demand. Ch. 13 in Gaudry, M. and R. R. Mayes, (eds). *Taking Stock of Air Liberalization*. Kluwer Academic Press, pages 189-222.

Mandel, B., Gaudry, M. and W. Rothengatter (1994). Linear or Nonlinear Utility Functions in Logit Models? The Impact on German High Speed Rail Demand Forecasts. *Transportation Research B* 28, 2, 91-101.

Mandel, B., Gaudry, M. and W. Rothengatter (1997). A Disaggregate Box-Cox Logit Mode Choice Model of Intercity Passenger Travel in Germany and its Implications for High Speed Rail Demand Forecasts. *The Annals of Regional Science* 31, 2, 99-120.

Mandel, B., Gaudry, M., Rothengatter, W., Jeiter, J., Last, J., Klar, R., Schnell, O., Daskaleas, T., and R. Masset (1991). *Modell zur Abschäzung der modalen Wirkungen von Investitionen im spurgeführten Hochgeschwindigkeits- Personenfernverkehr*. Forschungs- und Entwicklungsauftrag des Bundesministers für Verkehr FENr.: 602848/89. Institut für Wirtschaftspolitik und Wirschaftsforschung (IWW), Universität Karlsruhe (TH), 9/91.

Marshall, A. (1890). *Principles of Economics*, Macmillan and Co. Ltd, London.

McCarthy, P. S. (1982). Further evidence on the temporal stability of disaggregate travel demand models. *Transportation Research B* 16, 4, 263-278.

McCulloch, J. H. (1985). On Heteros*edasticity. *Econometrica* 53, 2, 483.

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. (1974). The measurement of urban travel demand. *Journal of Public Economics* 3, 303-328.

McFadden, D. (1975). *On independence, structure, and simultaneity in transportation demand analysis*. Working paper 7511, Institute of Transportation and Traffic Engineering, University of California, Berkeley, CA.

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

McFadden, D. (1976b). *The Multinomial Logit Model when the Population Contains « Captive » Subpopulations*. Unpublished Memorandum, 7 p., 30th September 1976.

McFadden, D. (1981). Econometric Models of Probabilistic Choice. In Manski, C. F. and D. McFadden (eds). *Structural Analysis of Discrete Data with Econometric Applications*, M.I.T. Press.

McFadden, D. (1998). Measuring Willingness-To-Pay for Transportation Improvements. Ch 15 in Gärling, T., Laitila, T. and K. Westin, (eds). *Theoretical Foundations of Travel Choice Modelling*, 339-364, Pergamon.

McFadden, D. and K. Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics* 15, 447-470.

McLynn, J. M., Goldman, A. J., Meyers, P. R. and R. H. Watkins (1968). Analysis of a Market Split Model. *Journal of Research of the National Bureau of Standards* 72-B, 1.

McLynn, J. M., and T. Woronka (1969). *A Family of Demand and Mode Split Models*. Arthur Young and Co., April.

Moïsi, F., Raison, M. and L. Silman (1981). A Mode-Split Model with Multiple Transit Modes. *Proceedings of the PTRC Summer Annual Meeting*.

Montmarquette, C. and A. Blais (1987). A Survey Measure of Risk Aversion. *Economics Letters* 25, 1, 27-30.

Montmarquette, C. and S. Mahseredjian (1985). Functional Forms and Educational Production Functions. *Economics Letters,* 19, 291-294.

Morgenstern, O. (1936). *Vorläufiger Bericht über einen Haupteil des Fragenkomplexes bezüglich des Wettbewerbes zwischen Eisenbahn und Auto in Oesterreich*. Oesterreichischen Institutes für Konjunkturforschung, 102 p., Vienna, 28 August. Forthcoming as Publication AJD-138, Agora Jules Dupuit, Université de Montréal, www.e-ajd.net.

Morrison, S. and C. Winston (1985). An econometric analysis of the demand for intercity passenger transportation. *Research in Transportation Economics* 2, 213-237.

Murthy, G. V. S. N. (1976). On the estimation of general functional form with random coefficients. *Sankhyā* C, 38, 37-43.

Nelder, J. A. and R. W. M. Wedderburn (1972). Generalized linear models. *Journal of the Royal Statistical Society A*, 135, 370-384.

Nelson, H. L. and C. W. J. Granger (1979). Experience with using the Box-Cox transformation when forecasting economic time series. *Journal of Econometrics* 10, 57-69.

Neumann, J. von and O. Morgenstern (1947). *Theory of Games and Economic Behavior*. Princeton University Press.

OQRTTF (1991). *Ontario/Quebec Rapid Train Task Force Final Report*. Ontario/Quebec Rapid Train Task Force, May.

Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* 70, 349, 120-126.

Orro, A., Novales, M. and F. G. Benitez (2005). *Nonlinearity and Taste Heterogeneity Influence on Discrete Choice Model Forecasts*. 18 p., Association for European Transport and contributors.

Orro, A., Novales, M. and F. G. Benitez (2010). Box-Cox mixed Logit model for travel behaviour analysis. Paper presented at the 8th International Conference of Numerical Analysis and Applied Mathematics, Rhodes, 4 pages, September. Forthcoming in *American Institute of Physics (AIP) Conference Proceedings*.

Ortúzar, J. de D. (2006). *Defining the State of the Art in Discrete Choice Modelling*. Paper presented at the Workshop on Transportation and Sustainable Cities, Universidad de Chile, Santiago, 15 pages, 7-11 August.

Oum, T. H. and D. W. Gillen (1983). The Structure of Intercity Travel Demand in Canada: Theory, Tests and Empirical Results. *Transportation Research B* 17, 3, pages 175-191.

Oum, T. H., Waters II, W. G. and J. S. Yong (1990). *A Survey of Recent Estimates of Price Elasticities for Demand for Transport*. WPS 359, Infrastructure and Urban Development Department. The World Bank, 22 pages, January.

Palma, A. de and O. Sanchez (1998). *Accessibilité dans un système de transport en environnement concurrentiel*. Working Paper 9804, Thema, 15 p.

Palma, A. de and J.-F. Thisse (1987). Les modèles de choix discrets. *Annales d'Économie et de Statistique* 9, 151-190.

Parra Granifo, R. S. (1995). Modelos de particion modal Logit Box-Cox para Santiago: interpretation microeconomica y valores del tiempo. In Martínez Concha, F. J. (ed). *Actas Septimo Congreso Chileno de Ingeniería de Transporte*, Santiago, pp. 169-186.

Park, R. E. (1966). Estimation with Heteroscedastic Error Terms. *Econometrica* 34, 4, p. 888.

Patterson, Z., Ewing, G. O. and M. Haider (2007). *Mode and carrier choice in the Quebec City -Windsor Corridor: a random parameters approach*. Paper presented at The 6th Triennal Symposium on Transportation Analysis (TRISTAN VI), 10-15 June 2007, Phuket Island, Thailand, 21 p.

Pearl, R. and L. J. Reed (1920). On the rate of growth of the population of the United States since 1790 and its mathematical representation. *Proceedings of the National Academy of Sciences* 6, 6, 275-288.

Pearl, R. and L. J. Reed (1929). The Population of an Area Around Chicago and the Logistic Curve. *Journal of the American Statistical Association* 24, 65, 66-67.

Pickrell, D. (1987). Models of intercity passenger travel demand. In J. Meyer and C. Oster (ed). *Airline deregulation and the New Airline Entrepreneurs*, MIT Press, Cambridge.

Pong, S. (1991). *L'application du modèle probabiliste et de la transformation de Box-Cox au choix du mode de transport au Chili*. Thèse de Maîtrise, Département de sciences économiques, Université de Montréal.

Pregibon, D. (1980). Goodness of link tests for generalized linear models. *Applied Statistics* 29, 15-24.

Quandt, R. E. (1970). *The Demand for Travel: Theory and Measurement*. Heath Lexington Books, Lexington, Massachusetts.

Quandt, R. E. and W. J. Baumol (1966). The Demand for Abstract Transport Modes: Theory and Measurement. *Journal of Regional Science*, 6, 2, 13-26.

Quarmby, D. A. (1967). Choice of travel mode for the journey to work: some findings. *Journal of Transport Economics and Policy* 1, 273-314.

Quiggin, J. (1979). *The Theory of the Multiproduct Firm Under Uncertainty*. B. ec. Honors thesis Australian National University.

Quiggin, J. (1981). Risk perception and risk aversion among Australian farmers. *Australian Journal of Agricultural Economics* 25(2), 160–9.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization*, 3, 323-343.

Rail Concept (2008). *Modèle Trafic TESE : Élasticité du trafic fer au tarif fer, V5*. Réseau Ferré de France, 6 p., Édition du 6 mars.

Ramjerdi, F. (1993). *Time in transport, theory and some empirical findings*. Institute of Transport Economics, TOI Report 178, 102 pages, Oslo.

Ramjerdi, F. and L. Rand (1992). *The national model system for private travel.* Institute of Transport Economics, TOI Report 150, 93 pages, Oslo.

Rassam, P. R., Ellis, R. H. and J. C. Bennett (1970). *The N-Dimensional Logit Model: Development and Application*. Peat Marwick and Mitchell & Co, Washington, D.C., 42 p.

Rassam, P. R., Ellis, R. H. and J. C. Bennett (1971). The n-Dimensional Logit Model: Development and Application. *Highway Research Record* 369, 135-147.

RATP and Cambridge Systematics (1982). *Études de politiques de transport en région Île-de-France. Mise au point et utilisation de modèles désagrégés de choix modal.* RATP, Direction des études générales, Service GL, 53ter Quai des Grands Augustins, Paris 6ème.

Rea, J. C., Platts, J. and M. J. Wills (1977). The Potential for Rail Passenger Services in the Windsor-Quebec City Corridor. *RTAC Newsletter*, Summer, 9-15.

Restle, F. (1961). *Psychology of Judgment and Choice*. John Wiley and Sons, New York.

Samuelson, L. (1985). On the Independence from Irrelevant Alternatives in Probabilistic Choice Models. *Journal of Economic Theory* 35, 2, 376-389.

Sävenstedt, G. and S. Uhlin (1985). Models of Swedish Interregional Passenger Travel. *Transportation Planning Methods,* PTRC, 13th Summer annual Meeting Proceedings, Vol. P-268, 89-106, July.

SC Stormont (1993). *High Speed Train Market Assessment Project*. Research commissioned by Air Canada and CP Rail System, SC Stormont Corporation.

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

Schneider, W. (2003). *The German National Freight Transport Model.* Paper presented at the Conference on National and International Freight Transport Models - With application to the Scandinavian – European Transport Corridor, Helsingør, September, 8 and 9, 11 p.

Schnetzler, B. (1996). *Modèles stochastiques de la décision. L'erreur de prédiction*. Rapport INRETS n° 207, Institut National de Recherche sur les Transports et leur Sécurité, 71 p., mars.

Schnetzler, B. (1998). *Modèles stochastiques du choix discret* - I et II. Unpublished mimeo manuscript, 70 p.

Schoch, M. (2000). *VACLAV-VIA, The IWW-MKmetric Passenger Model.* Institut für Wirtschaftspolitik und Wirtschaftsforschung, Universität Karlsruhe (TH), 20 p., September.

Schoch, M. (2003). *Methoden zur Bestimmung der geeigneten räumlichen Aggregation in Personenfernverkehrsmodellen*. Dr. rer. Pol. Dissertation, Fakultät für Wirtschaftswissenschaften der Universität Fridericiana Karlsruhe (Technische Hochschule), 151 p., Januar.

Seaks, T. G. and S. K. Layson (1983). Box-Cox estimation with standard econometric problems. *The Review of Economics and Statistics* 55, 1, 160-164.

Selz, T. (2004). Le modèle national d'Allemagne du traffic marchandises (*The BVU Nested Box-Cox Logit Freight Mode Choice Model Developed for the 2003 German Federal Transport Plan (FTIP))*. Presentation at the MODEV-M Workshop, Fondation Reille, Paris, 22nd June, 38 p.

Setec International and Stratec (2008). *Développement d'un modèle de prévision de trafic voyageur longue distance. Manuel de Référence*, Version 2. Réseau Ferré de France, 103 pages, Novembre.

Slutsky, E. E. (1915). Sulla teoria del bilancio del consumatore. *Giornale degli Economisti* 51, 1-26, July.

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

Smith, B., Abdoolakham, Z. and J. Taplin (2010). Demand and choice elasticities for a separable product group. *Economics Letters* 108, 134-136.

Solari, L. (1971). *Théorie des choix et fonctions de consommation semi-agrégées*. Librairie Droz, Genève.

Sonis, M. (1992). Universal analytical structures of discrete choice models. *Annals of Regional Science* 26, 349-360.

Spanos, A. (1987-88). Error Autocorrelation Revisited: The AR(1) Case. *Econometric Reviews* 6, 2, 285-294.

Spitzer, J. J. (1977). A simultaneous equations system of money demand and supply using generalized functional forms. *Journal of Econometrics* 5, 117-128.

Spitzer, J. J. (1982). A primer on Box-Cox estimation. *The Review of Economics and Statistics* 64, 307-313.

Spitzer, J. J. (1984). Variance estimates in models with the Box-Cox transformations: Implications for estimation and hypothesis testing. *Review of Economics and Statistics* 66, 645-652.

STIF (2004). *Prévisions de trafic régional sur l'Île-de-France; Fonctionalités des modèles – Méthodologie : Cas du STIF avec le Modèle ANTONIN*. Syndicat des Transports d'Île-de-France, 16 pages, June.

Stone, R. (1954). Linear expenditure systems and demand analysis: an application to the pattern of British demand. *Economic Journal* 64, 511-527.

Stopher, P. R. and J. O. Lavender (1977). Disaggregate behavioural demand models: empirical tests of three hypotheses. *Proceedings, Transportation Research Forum* 13, 32-36.

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

Swait, J. D. and M. Ben-Akiva (1987). Incorporating random constraints in dicrete choice models: an application to mode choice in São Paulo, Brazil. *Transportation Research B* 21, 2, 91-102.

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

Tardiff, T. J. (1978). *The Use of Alternative Specific Constants in Choice Modelling.* Working Paper UCI-ITS-SP-78-6, Institute of Transportation Studies, University of California, Irvine, 28 p., December.

Tegnér, G., Holmberg, I., Loncar-Lucassi, V. and C. Nilsson (2000). The DRAG-Stockholm-2 model. Ch. 5 in Gaudry, M. and S. Lassarre, (eds), *Structural Road Accident Models: The International DRAG Family*, Elsevier Science, Oxford. pp. 127-156

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

Tirole, J. (1988). *The Theory of Industrial Organization*, MIT Press.

Train, K. E. and D. McFadden (1978). The Goods/Leisure Trade-Off and Disaggregate Work Trip Mode Choice Models. *Transportation Research* 12, 5, 349-353.

Tran, C.-L. and M. Gaudry (2008a). *LEVEL: The L-1.6 program user guide with databases, examples and Tablex outputs*. Publication AJD-106, Agora Jules Dupuit, Université de Montréal, 354 p., September. www.e-ajd.net.

Tran, C.-L. and M. Gaudry (2008b). *LEVEL: The L-1.6 source and compiled programs, databases and examples.* Publication AJD-107, Agora Jules Dupuit, Université de Montréal, September 2008. www.e-ajd.net.

Tran, C.-L. and M. Gaudry (2008c). *LEVEL: The L-2.1 estimation procedures for BC-DAUHESEQ (Box-Cox Directed AUtoregressive HEteroskedastic Single EQuation) regression.* Publication AJD-108, Agora Jules Dupuit, Université de Montréal, 26 p., April, revised September. www.e-ajd.net.

Tran, C.-L. and M. Gaudry (2008d). *LEVEL: The L-2.1 program user guide with databases, examples and Tablex outputs*. Publication AJD-109, Agora Jules Dupuit, Université de Montréal, 147 p., November. www.e-ajd.net.

Tran, C.-L. and M. Gaudry (2008e). *LEVEL: The L-2.1 source and compiled programs, databases and examples*. Publication AJD-110, Agora Jules Dupuit, Université de Montréal, December. www.e-ajd.net.

Tran, C.-L. and M. Gaudry (2008f). *SHARE: The S-1 to S-5 estimation procedures for the Standard and Generalized BOX-COX LOGIT and DOGIT and for the Linear and Box-Tukey INVERSE POWER TRANSFORMATION-LOGIT models with aggregate data*. Publication AJD-114, Agora Jules Dupuit, Université de Montréal, 76 p., October. www.e-ajd.net.

Tran, C.-L. and M. Gaudry (2008g). *QDF: The Quasi-Direct Format calculation procedures to derive modal demand elasticities and diversion rates from products of coupled total demand and mode choice model results.* Publication AJD-117, Agora Jules Dupuit, Université de Montréal, 25 p., April, revised October. www.e-ajd.net.

Tran, C.-L. and M. Gaudry (2009a). *PROBABILITY: The P-2 to P-6 estimation procedures for the Standard and Generalized BOX-COX LOGIT and DOGIT and for the Linear and Box-Tukey INVERSE POWER TRANSFORMATION-LOGIT models with disaggregate data*. Publication AJD-111, Agora Jules Dupuit, Université de Montréal, 72 p., January. www.e-ajd.net.

Tran, C.-L. and M. Gaudry (2009b). *PROBABILITY: The P-2/P-6 program user guide with databases, examples and Tablex outputs.* Publication AJD-112, Agora Jules Dupuit, Université de Montréal, 218 p. April 2009. www.e-ajd.net.

Tran, C.-L. and M. Gaudry (2009c). *PROBABILITY: The P-2/P-6 source and compiled programs, databases and examples.* Publication AJD-113, Agora Jules Dupuit, Université de Montréal, April. www.e-ajd.net.

Tran, C.-L. and M. Gaudry (2009d). *SHARE: The S-1/S-5 Program user guide with databases, examples and Tablex outputs.* Publication AJD-115, Agora Jules Dupuit, Université de Montréal, 312 p., May. www.e-ajd.net.

Tran, C.-L. and M. Gaudry (2009e). *SHARE: The S-1/S-5 source and compiled programs, databases and examples.* Publication AJD-116, Agora Jules Dupuit, Université de Montréal, May. www.e-ajd.net.

Tran, C.-L. and M. Gaudry (2010a). ***QDF**: The **Q**uasi-**D**irect **F**ormat program user guide with databases, examples and TABQDF outputs.* Publication AJD-118, Agora Jules Dupuit, Université de Montréal, Forthcoming. www.e-ajd.net.

Tran, C.-L. and M. Gaudry (2010b). *QDF: The **Q**uasi-**D**irect **F**ormat source and compiled programs, databases and examples*. Publication AJD-119, Agora Jules Dupuit, Université de Montréal, Forthcoming. www.e-ajd.net.

Tran, C.-L., Gaudry, M. and M. G. Dagenais (2008). LEVEL: *The L-1.6 estimation procedures for BC-GAUHESEQ (**B**ox-**C**ox **G**eneralized **AU**toregressive **HE**teroskedastic **S**ingle **EQ**uation) regression and multimoment analysis*. Publication AJD-105, Agora Jules Dupuit, Université de Montréal, 59 p., April. www.e-ajd.net.

Transek AB (2003). *Utveckling av Sampers 2.1. Del 1 Estimeringen*. 143 pages, Solna, November.

Transport Canada (1979). SOMPS: *An Outline of PERAM Forecasts for Domestic Travel by Mode*. Draft preliminary working paper, Ottawa, Transport Canada, Strategic Planning Group.

Tukey, J. W. (1957). On the comparative anatomy of transformations. *Annals of Mathematical Statistics 28*, 602-632.

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

Veldman, S. J. and E. H. Bückmann (2003). A Model on Container Port Competition: An Application for the West European Container Hub-Ports. *Maritime Economics & Logistics* 5, 1, 3-22.

Verhulst, P.-F. (1838). Notice sur la loi que la population suit dans son accroissement. *Correspondance mathématique et physique*. Tome X, pp. 113-121.

Verhulst, P.-F. (1845). Recherches mathématiques sur la loi d'accroissement de la population. *Nouveaux mémoires de l'académie royale des sciences, des lettres et des beaux-arts de Belgique*, 18 :1–38.

Vrtic, M., Fröhlich, P., Schüssler, N., Axhausen, K. W., Lohse, D., Schiller, C and H. Teichert (2007). Two-dimensionally constrained disaggregate trip generation, distribution and mode choice model: Theory and application for a Swiss national model. *Transportation Research A*, 41, 857-873.

Walker, J. L. and Ben-Akiva, M. (2011). Advances in discrete choice: mixture models. Ch. 8 in Palma. A., Lindsey, R., Quinet. E. And Vickerman, R. (Eds.). *A Handbook of Transport Economics*, 160-187, Edward Elgar, 2011.

Wardman, M. (1997). Inter-urban rail demand elasticities and competition in Great Britain: evidence from direct demand models. *Transportation Research E* 33, 1, 15-28.

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

Wei, S.-J. (1996). *Intra-national versus international trade: how stubborn are nations in global integration?* Working Paper 5531, National Bureau of Economic Research, 34 pages.

Westley, J. D. (1979). The Demand for Urban Rail Transportation. *Journal of the American Statistical Association* 74, 576-587.

Wiener, N. (1956). The theory of prediction, Ch. 8 in E. F. Beckenback (ed.). *The Theory of Prediction*. McGraw-Hill, New York.

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.

Williams, H. C. W. L. and J. D. Ortuzar (1982). Behavioural Theories of Dispersion and Mis-specification of Travel Demand Models. *Transportation Research B* 16, 3, 167-219.

Wills, M. J. (1982). On the Equivalence of Common Rearrangements for Sum Constrained Travel Estimation Problems. *Transportation Research B*, 16B, 339-359.

Wills, M. J. (1986). A flexible gravity-opportunities model for trip distribution. *Transportation Research B*, 20B, 2, 89-111.

Wills, M. J., Rea, J. C. and F. B. Platts (1976). *Estimation and validation of a multimodal intercity travel demand model in Canada*. Transport Canada, Strategic Planning Group, 38 pages, September.

Yaari, M. (1987). The dual theory of choice under risk. *Econometrica*, 55, 95-115.

Young, K. H. (1969). An Abstract Mode Approach to the Demand for Travel. *Transportation Research* 3, 443-461.

Zarembka, P. (1974). Transformation of variables in econometrics. Ch. 3 in P. Zarembka (ed.). *Frontiers in econometrics.* p. 81-104, Academic Press, New York.