Improved Modelling of Competition among Airports through Flexible Form and Non Diagonal Demand Structures
Explaining Flows registered within a New Traffic Accounting Matrix

by

Marc Gaudry*

1 Directeur, Agora Jules Dupuit (AJD), www.ajd.umontreal.ca
Professeur, Département de sciences économiques
Université de Montréal, Montréal

2 Chercheur associé, Modélisation de l’Incertitude et du Risque des Transports (MIRT)
Département Économie et Sociologie des Transports (DEST), www.inrets.fr
Institut National de Recherche sur les Transports et leur Sécurité (INRETS), Arcueil

*The author (marc.gaudry@umontreal.ca) thanks Professor Jaap de Wit for comments on an early version of this paper invited for presentation at the Airmeth Workshop on Network strategies of multi-hub airlines and the implications for national aviation policies, The Netherlands Directorate-General for Civil Aviation and Freight Transport, The Hague, October 28, 2005. He also thanks Dr Benedikt Mandel of MKmetric GmbH, Karlsruhe, for his comments on this final version produced for the German Aviation Research Society (GARS) Workshop on Benchmarking of Airports and EU-Liberalization, Hamburg, February 23, 2006, and forthcoming in abbreviated form in Forsyth, P., Gillen, D., Müller, J., Niemeier, H.-M. and D. Starkie (eds), Competition in the European Airport Industry, Ashgate Burlington, 2008.

Université de Montréal
Agora Jules Dupuit, Publication AJD-121

Institut National de Recherche sur les Transports et leur Sécurité
Département économie et sociologie des transports, Working Paper DEST No 10

October 24, 2005; revised June 15, 2008
Abstract

Briefly stated, we first provide a new accounting framework that pinpoints the relevant flows to be explained by air demand and airport choice models. We then focus on the critical feature of demand modelling practice that bears particularly on airport competition and hub stability: the built-in IIA-axiom consistency of typical “diagonal” structural demand and itinerary choice procedures. We argue that this core property must be avoided because, always dubious, it is now particularly challenged in modelling the fastest growing component of passenger air demand, non-business trips, and by the clear need for reference alternatives in mode and path or airline company choice representation procedures. Because flows are interdependent, the utility of alternatives cannot be defined only by reference to own (matrix diagonal) transport conditions.

On this critical point, we summarize how Standard Box-Cox endogenous form specifications contribute to a much improved representation of the role of transport conditions within prevailing IIA-consistent structures but argue that freedom from “diagonal slavery” requires more, to wit: spatial correlation processes in Generation-Distribution models and Generalized Box-Cox specifications in Mode and Path choice analysis. In both cases, IIA-consistency is avoided in realistic ways by making parsimonious use of off-diagonal terms and permitting in principle the establishment of complementary alternatives, in contrast to the currently forced substitution straitjacket imposed on them.

In more detail, we first define a new four-part Traffic Accounting Matrix (TAM) to register all spatial flows of interest for air demand forecasting, effectively extending the scope of classical algebraic input-output analysis by doubling up and reinterpretating the intermediate and final transactions components of two-part Input-Output (IO) matrices. Strictly defined subsets of a TAM can then be matched to, and explained by, the usual procedures pertaining to the distinct generation, distribution, mode choice and assignment steps of traffic demand planning, or to their combinations.

We then focus on the key properties of such demand models in order to evaluate their relevance to the explanation of airport or hub competition and consider, among potential remedies, the estimation of form with Box-Cox transformations, but point out that their demonstrated relevance to the measurement of the impact of transport conditions is insufficient to solve the problem at hand. In both Generation-Distribution and Split Choice mode-company-path structural steps, the predominant use of Independence from Irrelevant Alternatives (IIA) consistent cores must be rejected to account properly for competition among destinations in Generation-Distribution models and for the prevailing importance of reference alternatives in Split Choice mode-company-path models.

We provide a first partial literature summary of numerous results obtained with endogenous functional forms in both of these structural steps but argue that, because the issue of non separability of utility is not directly addressed by standard Box-Cox transformations, their increased explanatory power and realism—as compared to the popular fixed form a priori logarithmic (in Gravity models) and linear (in Logit models) specifications— is more relevant to the proper measurement of the role of transport conditions (distance, level of service or price) than to the necessary representation of interdependence among alternatives, which mandates the abandonment of “diagonal slavery” in utility formulations.

Of course, the proper role of transport conditions still matters decisively in both Generation-Distribution models of transport or trade and in the Mode or company-path Choice splits. In the former class, proper curvature defines the total market reach and data determined forms rectify the demonstrably incorrect use of distance in the many logarithmic pooled time-series and cross-sectional models. In the latter class, allowing for changing marginal utility profoundly modifies the relative sensitivity of longer over shorter length trips, as compared to their behavior in prevailing untested linear constant marginal utility forms of the same functions, never theoretically very credible nor empirically sustainable. But none of these benefits and remedies to current dominant practice allows for interdependence (non separability) of utility, the key future demand modelling challenge if ex ante forecasts are to be of relevance to the air demand question at hand.

To point to real remedies, we summarize some recent promising attempts to deal with interdependence in manageable ways expected to yield “diagonal dominant” results: through the use of spatial autocorrelation in Box-Cox Generation-Distribution models and of Generalized Box-Cox specifications in Split models. Separable utility is thereby rejected by the data but without using too many independent off-diagonal terms pertaining to transport conditions: if the denial of any separability has been the scourge of classical demand equation system, its blind imposition has been that of Gravity and Logit demand systems. Considerate and flexible middle ways are now within reach and they matter most to model new interdependent markets, such as tourism.

Key words: traffic accounting matrix (TAM), input-output analysis, air traffic, O-D passengers, transfer passengers, Generation-Distribution, Gravity model, border effects, Mode Split, High Speed Rail, Path choice, Independence from Irrelevant Alternatives (IIA), non linearity of utility, non separability of utility, spatial autocorrelation, Box-Cox regression forecasting, Standard Box-Cox Logit, Generalized Box-Cox Logit, statistical correlation, regression sign, identification of Logit constants, Inverse Power Transformation-Logit, Canada, Germany, Quebec-Windsor corridor, European freight, Pyrénées, intermodal container train demand.
Table of contents

1. Introduction: models chasing reality ex post ................................................................. 1
2. Flow accounts, textbook demand equations and airport system components .................. 3
   2.1. Pinpointing flows identified within accounts and explained by typical equations ......... 3
   2.2. Airport demand, performance and supply components ............................................ 6
3. Consistency with IIA and structure of core equations determining hub flows .................. 10
   3.1. Two questions to be asked of each model component ............................................ 10
   3.2. The Generation-Distribution gravity-type part determining trip length .................... 11
       A. Within IIA-consistent structures: the issue of form ............................................. 11
       B. Away from IIA consistency: the issue of spatial competition ............................... 15
   3.3. The Split-type part determining mode/company/scheduled flight and path ............... 17
       A. Within IIA-consistent structures: the issue of form in Logit choice modelling ......... 18
       B. Away from IIA consistency in mode choice: diagonal dominance instead of diagonal slavery ................................................................. 24
       C. Implications for path, company or service choice in air markets .......................... 29
4. Conclusion: a TAM and remedies to build in Interdependence among Flows .................. 30
5. References .................................................................................................................. 32
6. Annex 1. The simple cost and frequency analytics of hubbing ........................................ 36

List of Tables

Table 1. Traffic accounting matrix (TAM) representation of the physical transport network flows ................................................. 4
Table 2. Flow matrix for the 4-city system without direct connections from Airport J to Airport B .............................................. 5
Table 3. Relationship between input-output (I-O) and traffic accounting matrices (TAM) ................................................................. 5
Table 4. Market and Network Analysis: a Three-Level Approach ........................................ 8
Table 5. Example of full price matrix used in three-mode representative utility function set ....... 11
Table 6. Flexible form, heteroskedasticity and spatial correlation in freight trade models ........... 12
Table 7. Flexible form, heteroskedasticity and spatial correlation in passenger transport models .............................................................................. 14
Table 8. Selected own share elasticities of Box-Cox Logit models used in the simulations .......... 23
Table 9. Comparing Linear with Standard and Generalized Box-Cox elasticities and values of time ................................................. 26
Table 10. Linear, Standard and Generalized Logit mode choice model of freight across the Pyrenees ....................................................... 27

List of Figures

Figure 1. Response asymmetry and non linearity: Classical Linear-Logit vs Standard Box-Cox-Logit ........................................................................................................ 19
Figure 2. Fare elasticities in intercity Box-Cox Logit model (Canada, 92 city-pairs, 1972) ........... 20
Figure 3. Travel time elasticities in intercity Box-Cox Logit model (Canada, 92 city-pairs, 1972) ........................................................................................................................................ 20
Figure 4. Comparing Linear and Box-Cox Logit forecasts of an ICE train scenario in Germany ......................................................... 24
Figure 5. Comparing Linear and Box-Cox Logit forecasts of European intermodal train gains ................................................................................................................. 24
Figure 6. Comparing Generalized Box-Cox and Linear Logit forecasts of intermodal train gains .................................................................................................................................................. 29
Figure 7. Two spatial configurations .................................................................................... 36
1. Introduction: models chasing reality \textit{ex post}¹

\textbf{Demand side question asked.} The identification of the determinants of airport hub demand, unavoidable in the recent context of the Air France-KLM integration treaty, is clearly of a general interest because competition among airports is widespread and measurable, as the detailed studies by Mandel (\textit{e.g.} 1999a, 1999b), unequalled to this day to capture airport market power, have shown for Germany. As our focus cannot be that of a survey of the literature on airport competition, we simply ask to which extent key features of current air demand modelling practice and its trends can be useful in understanding the general question of air market development simultaneously with that of the competition among airports. It is a challenge to address the issue of the structure of our modelling approach: our thinking can so easily prove inadequate.

\textbf{A context of supply-side thought market failure.} Indeed, when Alfred Kahn successfully proposed air deregulation in 1978, thereby ensuring that he would be the last chairman of the United States Civil Aeronautics Board (C.A.B.) abolished in 1984, nobody —not even Professor Kahn himself—, foresaw the future shape of the air route structure that would progressively ensue, either in the United States or internationally (Kahn, 2003).

This massive thought failure in forecasting performance by all academic², policy and business experts occurred despite the existence of the successful \textit{Federal Express} freight hub, initiated in 1973 by Fred Smith in Memphis, Tennessee, and clearly profitable since 1976. And some form of the same failure of vision occurred everywhere: even in perfectly hexagonal hub-shaped France, the beginning of the structuring of the Air France hub at Charles-de-Gaulle airport awaited 1996, a process not completed until perhaps 2003.

Of course, this industry is full of similar supply-side market forecasting failures: despite the exactly linear growth of Southwest Airlines over time since 1971—it has now been retiring pilots for about 7 years—, who forecasted the rise of the new stable low-cost carrier (LCC) business model firms challenging incumbent full-service airlines, at least on short high-density markets, clearly benefiting in the short run from the obvious accumulated excess supply of subsidized airports? Not even the supply siders...

\textbf{Hubber humility: making sense \textit{ex post}.} In retrospect, one can make some sense of route restructuring³, even if the process has yet to stabilize at the international level due to the continuing protective straightjacket of bilateral agreements. We all indeed know the \textit{ex post} arguments about the high frequencies and leg density economies induced by changes from gridding to hubbing (see Annex 1), and we are all aware that this supply side revolution may have meant, at least for a small minority of U.S. passengers (Morrison and Winston, 1999 or 2002), more circuitous routings, fortunately involving almost no interline connections replaced by on-line connections. And we all wonder how long small European capital cities can keep daily direct non-stop air service to North America in a liberalized environment. It also makes sense to many of us that, in the next phase of market evolution, rising income trends might then slowly increase the value of time and consequently the relative demand for direct passenger flights, if not also for jointly supplied freight services.

\textbf{Challenge: the workings of the demand side based on independent alternatives.} Despite difficulties, we take up the challenge and initiate our discussion of competition among airports by emphasizing their hub character and neglecting (except implicitly in the definitions of land-based access variables) the changes in

---

¹ The author (marc.gaudry@umontreal.ca) thanks Professor Jaap de Wit for comments on an early version of this paper invited for presentation at the Airmeth Workshop on \textit{Network strategies of multi-hub airlines and the implications for national aviation policies}, The Netherlands Directorate-General for Civil Aviation and Freight Transport, The Hague, October 28, 2005. He also thanks Dr. Benedikt Mandel of MKmetric GmbH, Karlsruhe, for his comments on this final version produced for the German Aviation Research Society (GARS) Workshop on \textit{Benchmarking of Airports and EU-Liberalization}, Hamburg, February 23, 2006.

² In 1993 interestingly, Mandel (1999a) estimated correct price curvatures of the demand curves but concentrated on issues raised by raising prices (the 50 DM Hamburg premium), not on the size of demand if one lowered prices with such curvatures which clearly could have correctly forecasted LCC demand. No one thought prices could fall so much.

³ Michael Tretheway has also argued many times since 2003 at Hamburg Airport Conferences that this movement was helped by incorrect accounting of multi-leg revenues by airlines, including some notable double counting of revenues.
competitive advantage that may arise from improvements in complementary\textsuperscript{4} land mode links. But our discussion needs a framework to pinpoint relevant flows and airport performance.

Consequently, we first provide a new accounting framework that pinpoints the relevant flows to be explained by air demand and airport choice models. We then proceed to isolate the crucial feature of demand modelling practice that bears particularly on airport competition and hub stability: the built-in IIA-axiom consistency of typical “diagonal” structural demand and path choice procedures.

**Structural remedy: introduce interdependence among alternatives.** We argue that this core property must be avoided because, always dubious, it is now particularly challenged by the fastest growing component of passenger air demand, non-business trips, and by the clear need for reference alternatives in mode and path or airline company choice procedures. Because flows are interdependent, the utility of alternatives cannot be defined only by reference to own (matrix diagonal) transport conditions.

**Endogenous form is necessary but insufficient.** On this critical point, we summarize how Standard Box-Cox endogenous form specifications contribute to a much improved representation of the role of transport conditions within prevailing IIA-consistent structures but argue that freedom from “diagonal slavery” requires more, to wit: spatial correlation processes in Generation-Distribution models and Generalized Box-Cox specifications in Mode and Path choice analysis. In both cases, IIA-consistency is avoided in realistic ways by making parsimonious use of off-diagonal terms and permitting in principle the establishment of complementary alternatives, in contrast to the currently forced substitution straightjacket imposed on independently specified alternatives.

\textsuperscript{4} For a formal discussion of microeconomic complementarity and substitution between air and land modes, notably high speed rail, see Gaudry (1998). Postal TGV trains in France show that complements occur also for many air services.
2. Flow accounts, textbook demand equations and airport system components

As just mentioned above, it is necessary to pinpoint the flows of interest. We do this by extending the standard input-output framework and recalling how demand forecasts cannot simply relate network flows to changes in “final demand” without taking due account of the spatial structure of flows and of the multiple and necessary ways of specifying technical coefficient matrices in transport, in contrast with the standard unique specification used in unspatialized inter-industry economics. We complement this set-up by explaining where the equilibrium levels of service come from.

2.1. Pinpointing flows identified within accounts and explained by typical equations

Key modelling developments and hub stability. If market reality speaks, can academic research say anything to increase the clarity of the message? I will select a few key new features of models and ask whether they are more likely to imply hub stability or not. In a formal sense, this question raises a problem of air company/air path choice but I will treat it without being very specific about issues such as detailed scheduling and flight coordination (Burghouwt and Wit, 2003) critical for hub performance, but within a representative schematic demand generating framework.

Two kinds of flows at a hub. We define our interest as that of the total flow of air passengers using a hub location (airport), assimilated to a city, and composed of both direct and circuitous elementary flows.

That total number, for a certain trip purpose \( g \), is the sum of the origin-destination (O-D) passengers, for whom the hub \( h \) is the origin or the destination city of the city pair considered, and of the transfer passengers, passing through \( h \) on their path between any O-D city pair:

\[
[Air \ flow \ through \ h] = [Direct \ O-D \ traffic \ between \ h \ and \ any \ j] + [Circuitous \ traffic \ through \ h], \quad (1)
\]

or

\[
T^g_{air, h} = T^g_{air, hj} + T^g_{air, ihj}, \quad i, j \neq h, \quad (2)
\]

where for simplicity all flows are assumed to be bi-directional between city \( h \) and any city \( j \) as well as between any city pair \( ij \) whose air passengers choose to go through hub city \( h \). Also, for similar reasons and without loss of generality, we associate the airport cities to the origin or destination zones.

To make our intents clear, Table 1 presents an accounting system to visualize the hub flows of interest in (2) before we use equations to explain them. Part A of Table 1 is a Traffic Accounting Matrix (TAM) reporting on four flows (D1, D2, C1 and C2) that enter Airport J of City 4 and follow either a direct or a circuitous path to three other cities, as indicated in part B. In this TAM, known O-D flows correspond to the bottom right hand side quadrant of Part A and passengers are registered when they enter airport J, when they use a flight, and when they leave an airport for the outside. One can apply such a TAM to a complete airport system such as that reported on in Table 2 for the same four cities listed in Table 1 but without the direct flight from Airport J to Airport B.

Note first that the system O-D matrix is asymmetric. Also, adding up registrations yields \( FE \), the vector of flows entering an airport from any outside zone, and \( NE \), the vector of flows entering an airport from any other airport. One can also compute \( NL \) and \( FL \), the corresponding vectors of flows leaving an airport for another airport or for outside zones, respectively. Combining these, the total entering flow vector is given by \( TE = FE + NE \) and the total leaving flow vector by:

\[
TL = NL + FL. \quad (3)
\]

The accounting identity of a TAM. As passengers do not accumulate in the airports, the accounting identity of the airport system states that the flow coming through an airport from the outside or from other airports must be equal to the flow leaving that airport for other airports or for the outside:

\[
TE' = TL \quad (4)
\]

5 In the remaining running text of Section 2.1, vectors are denoted in bold, but not matrices or scalars.
where the (\(^T\)) denotes the transpose operator. We emphasize that the accounting identity of the system is not that the entering and leaving vectors \(FE\) and \(FL\) are the same: in the Table 2 example, the entering and leaving vectors differ due to the asymmetry of the O-D matrix. But, of course, if \(e^T\) denotes the transpose of the unit vector, summing the flows entering all airports from the outside and the flows leaving all airports for the outside must yield the same total number of air trips for the purpose of interest \(g\), as expressed in:

\[
e^T FE^T = e^T FL^T = 2175 = T_{air}^g.
\] (5)

Table 1. Traffic accounting matrix (TAM) representation of the physical transport network flows

<table>
<thead>
<tr>
<th>Flow Matrix 1</th>
<th>Destination airport</th>
<th>Destination city</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(O-D)</td>
<td>B</td>
<td>T</td>
<td>J</td>
</tr>
<tr>
<td>Origin airport</td>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>C1</td>
<td>C2</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>D1+C1+C2+C2</td>
<td></td>
</tr>
<tr>
<td>Origin city</td>
<td>NE</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>FE</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TE</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A TAM is a doubled up input-output matrix. In such an accounting system, expounded further in Gaudry (1973 or 2007), it is possible to pursue the obvious analogies with economic input-output (I-O) tables made up of N and L sub-matrices (extended with the lower-level Entering and Origin-Destination matrices in Table 3) and show for instance the general lack of proportionality between forecasts of the leaving demand vector \(FL^*\) and forecasts of network leg flows among airports \(N^*\). This is done by defining first the familiar matrices of direct (and, later, of indirect) input-output coefficients of inter-industry economics now available in two potential formats instead of one, as can be demonstrated with the help of Table 3 where a TAM is written in simplified format.

Two matrices of technical coefficients can also be defined. There are two possible matrices of technical coefficients because, after an explicit distinction is made between entering and leaving demand structures E and L, there are two ways to define the operation required to compute direct technological coefficients. Network flows N can be combined either with the leaving flows or with the entering flows; this possibility is
absent from input-output tables (Leontief, 1941) that effectively contain only N and L matrices\(^6\) and thereby impose a unique definition of the matrix of technological coefficients A, namely, in the familiar way:

\[ A = N \cdot B^{-1} \text{ or } N = A \cdot B, \]  

where B is a diagonal matrix with elements made up of components of the leaving demand vector TL:

\[ B \equiv \begin{bmatrix} \text{TL}_1 & \text{TL}_2 & \text{TL}_3 & \text{TL}_4 \end{bmatrix} \]

\(\text{Table 2. Flow matrix for the 4-city system without direct connections from Airport J to Airport B}\)

<table>
<thead>
<tr>
<th>Origin airport</th>
<th>Destination city</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>150</td>
<td>500</td>
</tr>
<tr>
<td>T</td>
<td>250</td>
<td>200</td>
</tr>
<tr>
<td>M</td>
<td>300</td>
<td>200</td>
</tr>
<tr>
<td>J</td>
<td>25</td>
<td>400</td>
</tr>
<tr>
<td>NE</td>
<td>250</td>
<td>2000</td>
</tr>
</tbody>
</table>

\(\text{Table 3. Relationship between input-output (I-O) and traffic accounting matrices (TAM)}\)

\[ A = N \cdot B^{-1}, \text{ or } N = A \cdot B, \]  

\(\text{Although, after some hard work, Quesnay’s open economy Tableau Oeconomique (1758) “balanced” expenditures among 4 population sub-segments (farmers, farmhands, landlords and artisans), it is easier to rewrite it as a 5-sector Leontief transactions table in order to make the foreign sector explicit (Brems, 1986). Crucially, Quesnay did not distinguish between intermediate and final flows as did Leontief by specifying N and L parts while keeping the N matrix square for inversion purposes. After Blankmeyer (1971), we simply interpret L as reporting on flows « towards the outside », which naturally calls for the « from the outside » counterpart matrix E and for their O-D combination in the remaining quadrant of Table 3. This new spatial extension differs deeply from the standard spatialization of Leontief systems expounded by Moses (1955) and adopted in transport (e.g. Cascetta and Di Gangi (1996), where a diagonal spatial trade matrix T is used to multiply a redefined diagonalized matrix A and suitably redefined output vectors).} \)
so that:

\[ \mathbf{NL} = \mathbf{A} \cdot \mathbf{TL}, \]  

and the usual operations can be effected, first by substituting (7) into (3), and so on.\(^7\)

**Matching TAM components to traffic modelling steps.** Independently from its extended input-output interpretation, the TAM found in Table 2 can be used to focus visually on the different objects to be explained by particular transport models. For instance, the Trip Generation step of the four-step transport planning format involves explaining the vectors \( \mathbf{T}_* \) and \( \mathbf{T}'_* \), noted there in outlined symbols. Our interest is in Generation-Distribution models (matrix O-D) and in Split mode/company/path choice models (matrix N).

In this analysis, we first match cities to airports, which imposes a particular structure of the network entering and leaving vectors \( \mathbf{NE} \) and \( \mathbf{NL} \) because (without loss of generality) flows are not allowed to enter or leave an airport from or towards more than a single zone. We then isolate the hub values of interest by underlining them or putting them in bold in Table 2. The underlined entries circumscribe the O-D passenger flows and isolate them from the transfer passenger flows. These are the flows to be explained by econometric models.

**A reference equation structure to explain flows of interest.** We explain them by associating representative analytical expressions to the two kinds of flows using hub \( h \), and consequently explicate (2) as follows:

\[
T_{air, h}^\gamma = \left\{ \text{Pop}_h^* \cdot \sum_j \left\{ \text{Pop}_j^* \cdot [U_{ij}]^\alpha \cdot [U_{air, ij} / \sum_m U_{m,ij}] \right\} \right\}, \quad j = 1, \ldots, Z, \quad (8)
\]

\[
+ \sum_i \sum_j \left\{ \text{Pop}_i^* \cdot \text{Pop}_j^* \cdot [U_{ij}]^\beta \cdot [U_{air, ij} / \sum_{m,ij} U_{m,ij}] \right\} \cdot [U_{air, ihj} / \sum_{air, ikj} U_{air, ikj}], \quad i, j \neq h, \quad (9)
\]

where (8) refers to O-D passengers and (9) to transfer passengers, \( \alpha \) and \( \beta \) are parameters, there are modes \( m = 1, \ldots, M \), and zones \( i, h, j = 1, \ldots, Z \), and:

- \( \text{Pop}_j^* \): income and activity weighted measures of population at location \( j \);
- \( [U_{ij}] \): the aggregate utility of the modes from \( i \) to \( j \) built from the denominator of:
- \( [U_{air, ij} / \sum_{m,ij} U_{m,ij}] \): the air modal market share from \( i \) to \( j \);
- \( [U_{air, ihj} / \sum_{air, ikj} U_{air, ikj}] \): the share of air trips from \( i \) to \( j \) going through hub \( h \).

Note that a possible definition of a successful hub is a location where the ratio of (9) to (8) is high. Note also that the quality of airport services is hidden in the definition of the utility terms consisting in “supply side” level of service or price variables found in modal utility functions defined in more detail below. Although we will not model the supply side, it is useful\(^8\) to make clear how the levels of variables used in the demand models arise.

**2.2. Airport demand, performance and supply components**

This “supply side” notion requires development because it is insufficient without the further introduction of the notion of performance, a notion of great practical import in the determination of air service levels and, in consequence, in the explanation of competition among hubs. Why and how?

**The 19th Century: from one to two levels.** Until the end of the 18th Century, explanations tended to relate variables of interest, say imports or exports flows, to hypothesized causes, whatever these may have been: money supply, the price level, etc.

\(^7\) From \( \mathbf{TL} = \mathbf{A} \cdot \mathbf{TL} + \mathbf{FL} \), one collects terms and writes \( \mathbf{TL} = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{FL} \) thereby deriving \( \mathbf{C} = (\mathbf{I} - \mathbf{A})^{-1} \), the famous matrix of indirect coefficients. Forecasting \( \mathbf{N}^* \) involves the reverse path from \( \mathbf{FL}^* \) to \( \mathbf{TL}^* \) and then to \( \mathbf{N}^* \) from \( \mathbf{N}^* = \mathbf{A} \cdot \mathbf{B}^* \) in (6).

\(^8\) As suggested by Professor Jaap de Wit.
Implicitly, single-layer (i.e. single equation) systems were effectively in use. Not surprisingly, the expected directions of the various effects of “causes” were often confusedly expressed, as compared to what will be the case when observations on quantities or flows will be thought of as resulting from the interaction of supply and demand. It will then be clear that many “causes” jointly affect supply and demand and have effects of very uncertain direction on equilibrium, effects often consisting in mixtures of underlying structural effects of diverse strength and expected sign.

Groenewegen (1987) reminds us that the phrase “supply and demand” was initiated in the context of price determination by Steuart-Denham (1767) and occurred infrequently until Ricardo (1817) used it in a chapter heading. Until the 1830’s, the terms were rarely used in the modern sense, i.e. each as a function of price: Cournot (1838) was the first to give such a systematic and symmetrical exposition of two sided markets. As these structural relationships were fleshed out, two structural equations, one for Supply and one for Demand (and an equilibrium condition), were deemed essential to the explanation of market data on quantities and prices in most markets.

The clarity of this structural mechanism progressively made it self-evident to all who read Cournot or, 4 to 5 years later, Dupuit (1844). But the empirical conditions under which one could “identify” each equation were not obvious until Working (1927) pointed out that their unscrambling (“identification”) from the data was possible for each equation only as long the other moved independently: if you imagine that the supply schedule is fixed and that the demand schedule moves by itself, clearly observed points will draw the supply curve, and conversely… A demanding task, identification is often assumed away in the rush to the computer.

The 20th century: from two to three levels in transportation. For us, however, the simplest way of thinking of transportation, and perhaps also of many other economic systems, is not to adopt this two-level Demand-Supply formulation: it is to add a third level, the determination of Performance that depends on both Demand and Supply, as researchers effectively do in structural transport analysis.

Some years ago, we introduced (Gaudry, 1976, 1979) this three-layer structure to capture the fact that realized transportation service levels often differ from supplied service levels and are best modeled through a third and explicit level between the classical supply and demand levels. At first, we first called the resulting structure « Demand-Cost-Supply » (D-C-S) to distinguish it from the classical « Demand-Supply » (D-S) pair. In our new structure, “costs” denoted realized money, time, crowding or safety levels. We also estimated with monthly time-series data a complete simultaneous three-layer bi-modal urban model system on these lines (Gaudry, 1980), within the system definition schematized in Table 4, to which we now turn because it is essential to understand airport performance elements that enter the utility functions.

Some conceptual implications of a three-layer system. Naturally, using a D-C-S system instead of the classical D-S system gave rise to unheard-of equilibria, such as the new « Demand-Generalized Cost » equilibrium, now distinct from the « Demand-Supply » equilibrium defined within the same 3-layer system.

To make the enriched formulation more accessible within the wide transportation planning subculture, we subsequently re-labeled the triplet D-C-S as D-P-S (Demand-Performance-Supply) and changed the notation (Florian and Gaudry, 1980, 1983) to that used in Table 4. At the time, the Supply side vector typically contained road infrastructure capacity and transit vehicle service characteristics pertaining to congested urban network simultaneously used by fixed route public and flexible route private vehicles. But it is general.

In the actual format of Table 4, the achieved Performance [P, C] contains realized queues, levels of congestion and risk, as well as other forms of flexible modal bearing ratios (effective capacity, occupancy or load factor and crowding, etc.) conditional on both actual Demand D and given Supply actions [S, T, F]. For instance, in a network equilibrium, there is a set of values of P, C and D that simultaneously satisfy the demand functions and (for given supply) the conditions required by the performance procedures. For our purposes here, money and time performance by origin-destination pair on the network have to be consistent with the demands generated under these network conditions, a non-trivial problem because the dimensions of the demand functions (from i to j) are not the same as those of the transportation performance conditions on individual network links a. And of course demand depends on the whole of the performance vector elements.
Other three-layer systems. This three-layer specification applies well to many regulated markets, notably those of communist economies (Gaudry and Kowalski, 1990), or to sectors of market economies (such as health) where the prices and freely determined wait times are not allowed to clear the markets but are “centrally planned” and regulated. And, to the extent that wait time partly defines quantity supplied (and the resulting logistics chain) even in unregulated markets, explicit performance modelling has a future.

Note in passing that modeling the three levels in this way to explain observations generated in the absence of D-S equilibrium is much simpler than using disequilibrium econometrics—a difficult combinatorial game—or some subtle forms of hysteresis. As an example of the former, Portes et al. (1987) stunningly concluded that the Polish economy was in «excess supply half of the time during the 1960’s and during the years 1976-1978!» It would have seemed more appropriate to have built a model in which the length of the queues for housing, cars, etc. was explained within a Performance level, as we do in transport, and to have raised in this context issues of identification the neglect of which produced such a “supply of revelations on centrally planned economies!” (Podkaminer, 1989).

Table 4. Market and Network Analysis: a Three-Level Approach

<table>
<thead>
<tr>
<th>D = Dem (P, C, Y, A)</th>
<th>DEMAND PROCEDURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>[P, C] = Per (D, [S, T, F])</td>
<td>PERFORMANCE PROCEDURE</td>
</tr>
<tr>
<td>[S, T, F] = Sup (SO, RE, [(W(S*, T*)], ST)</td>
<td>SUPPLY ACTIONS PROCEDURE</td>
</tr>
</tbody>
</table>

with:
- D: market demand
- P: out-of-pocket unit expenditures
- C: levels of service
- Y: consumer socio-economic characteristics and their budget
- A: economic activity
- S: quantity supplied
- SO: supplier objectives
- T: scheduled service levels
- RE: regulatory environment
- F: scheduled price, or fare
- ST: suppliers’ estimate of the state of the system
- [W(.)]: set of minimum cost combinations for the realization of any scheduled (S*, T*)

and where: D**, P** and C** denote realized values of demand, unit costs and service levels.

Most recent version: Gaudry (1999).

Similarly, the explicit modeling of Performance (queues, safety, etc.) should take precedence over the sometime fashionable search for hysteresis in various markets, such as that for labor of retail gasoline. And, in health studies, one should not conclude that reducing the number of doctors reduces public health expenditures but rather look at the length of queues, at the black market and at the market for side-privileges, and model these explicitly as the supply of doctors is manipulated. Of course, modelling Performance, as transport researchers do, is hard work. For instance, it does not suffice to model the stationarity of socialist economy shortages (Kornai, 1982): “insatiable demand” does not exist, but queues (Kornai and Weibull, 1977) and black markets do, and they reestablish equilibrium.

Airport performance and travel time. In consequence, it is important for our understanding of the airport system problem to distinguish among the three layers because Demand, Performance and Supply jointly determine airport passenger and freight transfer and delay times that are themselves critical objects of analysis along with “demand” and “supply” quantities. Airport performance elements partly define levels of service used in the representative demand functions that we presently concentrate on: e.g. the coordination of
flight required for hub-and-spoke networks creates congestion peaks (Burghouwt and Wit, 2003). Paradoxically, such time nuisances are immediately internalized because the peaking involves planes of the same company: higher peaking reduces the need for congestion charges, if not for terminal use charges!
3. Consistency with IIA and structure of core equations determining hub flows

Our analysis of econometric demand models generating hub demands proceeds in three parts. First, we state the two key questions to be asked about any demand model: whether it was estimated with fixed or variable mathematical forms and whether it allows for substitution and complementarity among alternatives. Second, we analyze Generation-Distribution models in current use from that perspective: it is not possible to discuss continental airport demand sensibly without a structure that determines total flows by all modes. Thirdly, we analyze the Mode/Path choice component of models from the same double-barrel perspective: this step is important not only to explain continental air traffic flows and itinerary choices but to model intercontinental air company and path choices, notably those that matter for competing hub demands.

3.1. Two questions to be asked of each model component

Two coupled model parts. Each of (8) or (9) can usefully be broken down between a Generation-Distribution part explaining total flow (by all modes or passenger paths, depending on the case) and a Split choice term explaining shares of alternatives (a mode, company or air itinerary, depending on the case). We note that the two parts are coupled through a total utility term $U_{ij}$ which we assume here (keeping formulation (9) simple at the cost of incompleteness) to be built from the denominator of the split term (a mode, company/path or combined hierarchical nested Logit choice model). In consequence, each part contains transport conditions and other, typically socio-economic, variables but we focus on the former.

For each model part, two questions. In each part, we consider two principal dimensions of import for us. The first one is that of the use of Box and Cox (1964) transformations of any $X_k$, written commonly as:

$$
X_k^{i(i)} = \begin{cases} 
\frac{X_k^{i(u)} - 1}{\lambda_i} & \text{if } \lambda_i \neq 0, \\
\ln X_k & \text{if } \lambda_i \to 0.
\end{cases}
$$

This transformation is of interest due to the numerous model improvements obtained through its use during the last 30 years. Of course, gains in theoretical reasonableness and fit are clear. But, as signs in regression depend not just on the variance of regressors considered individually but also on their covariances, and as these are modified by transformations of variables, the very existence of statistical correlation depends on mathematical form (and conversely, as form parameters are jointly estimated). The first question asked of models is therefore whether their form was assumed or estimated because it is easy to argue that behavioural reasonableness, numerical best fit and statistical robustness of coefficient estimates require rejection of untested results based on a priori fixed forms.

Note in passing that the use of Box-Cox transformations pertains to a fundamental problem about the existence of statistical correlation (causality). David Hume believed that we get the idea of 'necessary connection' from the constant conjunction of causes and effects: our point is that curvature and covariances are jointly determined and consequently jointly decide on the presence of 'constant conjunction' which cannot be reliably determined on the basis of a priori form.

The second dimension can be best defined if we start from the traditional specification of complete demand equation systems where, using Logit terminology for clarity, any utility function $V_m$ of an alternative can in principle include the transport conditions (prices or service levels) of all other alternatives, as in the land freight example of Table 5 where such conditions (prices in this case) are laid out in square matrix format (each accompanied by a regression and a form parameter, both here subjected to some equality restrictions).

Of course, traditional diagonal dominance, defined as strong own elasticities and weak cross elasticities, is expected in demand systems defined by the Table 5 format, but diagonal slavery is something else: when the cross terms are absent from (12)-(13), IIA-consistency is per force imposed (Luce, 1959, 1977). The second question asked of models is therefore whether such substitution or complementarity “cross” terms have been considered or not.
Concerning interdependency among markets, we will argue that it is one thing to estimate models ex post using IIA-consistent formulations, but that more flexible formulations should be used to account for interdependency, and that the importance of this generalization arises notably through tourism for Generation-Distribution models and more generally in mode or path choice (Split) models where a reference alternative matters to the explanation of consumer choices.

This implicitly means that ex ante airport demand forecasts are all the more biased that incorrect form and real interdependency are present but have not been duly accounted for. Our examples are drawn from passenger and freight markets and their implications for airport demand are briefly stated.

3.2. The Generation-Distribution gravity-type part determining trip length

Transport trip length is increasing rapidly: in France for instance, average distance to work doubled from 6 to 12 km from 1993 until 2001 and air trip length doubled to 1200 km between 1981-82 and 1993-1994 and is expected to have increased further in the current survey (2006-2007). During the 1980-1995 period, world air ton-km increased 20% more than air tons. Total market size (for all modes) and trip length are central to any airport service demand issue.

A. Within IIA-consistent structures: the issue of form

Gravity models and the IIA axiom. The gravity structure within (8) and (9) is arguably the most successful modelling structure in all of economics, compatible with many theories (e.g. Rossi-Hansberg, 2005), and still formidably robust empirically in both mathematical form and broad specification.

Suggested in strict Newtonian form by Carey (1858) to describe human spatial interactions, it was first used in transportation by Lill (1891) on Vienna-Brünn-Prague passenger rail flow data. We write it here in so-called « Generation-Distribution » form, without the double constraints that guarantee that all of what comes out of i or comes into j is assigned. We exclude such constraints because they bias the estimation of parameters by imposing a very special restriction on error terms and thereby artificially induce a form of accounting substitution among all O-D pairs that is absent from the core behavioural model structure. Indeed, this basic unconstrained Gravity structure may be written simply:

\[(Flow\ indicator_{ij}) \leftarrow (Activities_{i\ and\ j \ ;\ Socioeconomic\ i\ and\ j \ ;\ Ease\ of\ interaction\ ij), \quad (15)\]

where the ease, or (dis-)utility, of interaction is an impedance metric that can be of many types (generalized cost, geographical, preference or taste, cultural, legal or regulatory, etc.). This specification makes abundantly clear the fact that, because the flow from i to j depends only on variables that have i and j indices, the characteristics of other paths or locations have no influence on the flow from i to j. Consequently, the ratio of the demands for two O-D pairs does not depend on the characteristics of other links or locations than those involved in the ratio, a property consistent with Luce’s Independence from irrelevant Alternatives (IIA) axiom of choice theory. We wish to address the relevance of both mathematical form improvements, first studied on intercity trips in Georgia by Kau and Sirmans (1979) who, contrary to most other authors have since found, could not reject the multiplicative gravity form, and of IIA axiom avoidance efforts within this structure.

Form and response to transport conditions. In (15), trip length depends on the response to the “ease of interaction”, i.e. to transport conditions which are strongly correlated with distances in any spatial cross-section. Actual response depends on the true shape of the demand curve: we will note that, as a rule, its elasticity is not constant and that all untested constant elasticity log-log models generate incorrect trip length forecasts and incorrect diagnostics of the changing role over time of distance, used as a proxy for cost.
Start with freight trade models, all still using distance and all less advanced today than intercity transport models were 30 years ago with both flexible form and rich border effect specifications. In Table 6, using Box-Cox transformations on known freight models greatly increases log likelihood (comparing columns 2 to 1 and 6 to 5), despite estimated Box-Cox power values differing little numerically from 0, and shows that the optimal form flow elasticities with respect to distance, evaluated at sample means, are almost identical across continents at about -1,20 [in detail, -1,21 and -1,13 without heteroskedasticity and -1,16 and -1,25 with it]. Clearly, one would not wish to use the fixed form constant elasticity results, except as a first cut.

**Table 6. Flexible form, heteroskedasticity and spatial correlation in freight trade models**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>YEARLY FREIGHT FLOWS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaudry et al., (1996)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>McCallum (1995)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Variable</strong></td>
<td>Column Case 1</td>
<td>Column Case 2</td>
</tr>
<tr>
<td></td>
<td><strong>1</strong></td>
<td><strong>2</strong></td>
</tr>
<tr>
<td></td>
<td>Log-log</td>
<td>Box-Cox</td>
</tr>
<tr>
<td>Distance</td>
<td>elasticity</td>
<td>-1,42</td>
</tr>
<tr>
<td></td>
<td>t-statistic</td>
<td>(-22,66)</td>
</tr>
<tr>
<td></td>
<td>Box-Cox λ</td>
<td>0,00</td>
</tr>
<tr>
<td>Border dummy</td>
<td>elasticity</td>
<td>3,09</td>
</tr>
<tr>
<td></td>
<td>t-statistic</td>
<td>(23,69)</td>
</tr>
<tr>
<td>Other variables</td>
<td>Spatial autocorrelation ρ</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>Distributed lag π</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>Log-likelihood</td>
<td>-8089</td>
</tr>
<tr>
<td></td>
<td>Degrees of freedom</td>
<td>0</td>
</tr>
</tbody>
</table>

**Distance often an unacceptable surrogate for changing transport conditions.** But how good is distance at representing true transport conditions? This matters if the simplistic log-log form of Columns 1 and 5 is used to compare impedance effects from samples pertaining to different moments of time (Buch et al., 2004): assumed proportionate changes in costs between the samples will be accounted for in the constant but not 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 those of mere fit — some “distance” effects can only be identified if the log-log form is not optimal!

Conversely, the use of distance as a proxy for price in time series, say from 1970 until 1995 or 1999 (Ghosh and Yamarik, 2004; Fratianni and Kang, 2006) is incorrect in a gravity trade model of log-log form, even with period dummy variables: it simply reflects basic constant propensities to trade with more or less distant partners, independently from the desired critical role of changing transport conditions in explaining increased globalization flows. Therefore distance has to be used in addition to transport costs if trade flow explanations are tested with time-series data, as some authors do (Giuliano, 2006). And, to the extent that more sophisticated measures of impedance than distance are in rough proportion to it, any proper comparison of

---

9 For instance Gaudry and Wills (1978) demonstrate both the effect of flexible forms on elasticity estimates and the import of (non-dummy) linguistic border effects within Canada in 1972. Concerning the former, the optimal value of the λ used for the Uij utility term is 0.03 but the log likelihood gains of 22 units of log likelihood (4 degrees of freedom are involved due to the fact that all other variables are also optimally transformed) are massive even if the estimated value differs numerically only slightly from 0. For freight, even regional studies use border effects: in their study of 1989 France-Germany freight flows among 36 regions, Becker et al. (1994) show that the border divides total freight flows by 8,25 (with corresponding modal values of 25 for train, 6,78 for road and 0,33 for rivers).

10 Results for North America shown in B were presented to McCallum in June 1994 to check his previous results (McCallum, 1993) before their publication as A, but he chose the fixed form model results of Column 1, as did Helliwell (1998) after seeing B results for both North America and Europe in February 2006, despite their quality of fit.
the roles of distance and of its more sophisticated replacements, even for the same period, also requires that the form not be logarithmic.

**Impedance: from distance to a total utility term U.** To answer the question of the quality of distance used as a proxy for transport conditions, turn to the advanced intercity passenger models accounting for “induction” through the sophisticated $U_0$ utility term in (8) or (9) and using Box-Cox transformations to obtain an adequate fit. In all models of Table 7 the log-log form is demonstrably less adequate than the Box-Cox form.

In Parts II and III, distance is replaced by utility derived from the sum of the mode choice denominator terms: obviously, apart from sign which must be reversed, the distance and utility elasticities (indicated in bold on grey backgrounds) are extremely close. Certainly, any sophisticated index of the structure of modal transport conditions (such as the particular Expected Maximum Utility index consisting in the (log of) the denominator of a Logit model) is strongly correlated with distance in any spatial cross-section. Although optimal forms of the impedance variable are not too far, numerically speaking, from the log form of the initial popular gravity model, it makes a difference to use the best fitting form even in a cross section.

**Form and results for other regressors.** We note in Table 6 that the border effects decrease by about 10% if Box-Cox forms are used (comparing columns 2 to 1 and 6 to 5). The German data base is homogeneous and was not obtained like the Canadian one by merging heterogeneous sources (one providing spatial structure and another providing totals): its border dummy variable is therefore interpretable without undue\(^{11}\) risk.

Concerning other variables, in particular activity terms, it is extremely difficult to develop Generation-Distribution models of international passenger flows by inclusion of activity variables other than population, perhaps corrected by income and some measure of employment structure. We do not see any new modelling trends in this area, except the innovation introduced by Last (1998) who successfully weighted the trips by the propensity to travel by age class. But, a form of aggregation of market segments, this innovation has no definable effect on the structure of the hubs markets. There have also been a number of attempts to use Box-Cox transformations to estimate something more general than a simple multiplicative form of activity variables. Although exponents other than 1 are found and differ as between emission and attraction zones, the gains beyond fit do not directly bear on our issue except to weigh heavily against attempts to use system constraints to modify the unconstrained basic form of the Gravity equation.

\(^{11}\) Both McCallum, co-author of B, and Helliwell used only the North America database (supplied to us by McCallum in 1994 to test his specification) containing a mixture of data from two sources within Statistics Canada [whereby errors of scaling, bi-proportional Fratar (1954) or not, of interregional flows would directly modify only the USA-Canada border dummy variable]. The German database, also supplied to us by Blum in 1994 to produce the intercontinental comparison found in B, does not suffer from attempts to merge distinct databases on trade total and spatial structure: its homogeneity makes the border dummy variable meaningful and its usual interpretation admissible. Fortunately, the measurement of the role of impedance in the determination of trip length is unaffected by the merging error in the North America data set and straightforward with the European data set. We carried out tests with the data bases as supplied.
Table 7. Flexible form, heteroskedasticity and spatial correlation in passenger transport models

<table>
<thead>
<tr>
<th>Part I</th>
<th>Single-country models, all modes and all trip purposes (domestic flows)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaudry et al., (1992)</td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td>Column Case</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Utility elasticity</td>
<td>0.63</td>
</tr>
<tr>
<td>t-statistic</td>
<td>(42.33)</td>
</tr>
<tr>
<td>Other variables</td>
<td></td>
</tr>
<tr>
<td>Box-Cox λ</td>
<td>0.00</td>
</tr>
<tr>
<td>Spatial autocorrelation ρ</td>
<td>0.00</td>
</tr>
<tr>
<td>Distributed lag π</td>
<td>-</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-1318</td>
</tr>
<tr>
<td>Degrees of freedom</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Part II</th>
<th>Joint France (1993-1994) and United Kingdom (1991) model, all modes by trip purpose (domestic flows and flows between them or with 8 other European countries)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Last (1998)</td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td>Column Case</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Distance elasticity</td>
<td>-0.23</td>
</tr>
<tr>
<td>t-statistic</td>
<td>(-8.71)</td>
</tr>
<tr>
<td>Utility elasticity</td>
<td></td>
</tr>
<tr>
<td>t-statistic</td>
<td></td>
</tr>
<tr>
<td>Other variables</td>
<td></td>
</tr>
<tr>
<td>Box-Cox λ</td>
<td>0.00</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-65407</td>
</tr>
<tr>
<td>Degrees of freedom</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Part III</th>
<th>Business trips full set of 5711 observations</th>
<th>Business trips subset of 544 observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaudry et al. &amp; Last (1998)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td>Column Case</td>
<td>13</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Distance elasticity</td>
<td>-0.33</td>
<td>0.20</td>
</tr>
<tr>
<td>t-statistic</td>
<td>(-8.78)</td>
<td>(-11.02)</td>
</tr>
<tr>
<td>Utility elasticity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-statistic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Box-Cox λ</td>
<td>0.00</td>
<td>-1.02</td>
</tr>
<tr>
<td>Spatial autocorrelation ρ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-50846</td>
<td>-50762</td>
</tr>
<tr>
<td>Degrees of freedom</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

---

12 Results drawn from Table B.2, Columns 1, 3 and 5. The Utility term is constructed from a 4-mode Box-Cox Logit model also found in Liem and Gaudry (1999) and specified in RATE format (for this, see a discussion below) with distinct transformations applied to its four network variables (Price, Speed, Distance, Frequency) but not its Socio-economic variables.

13 Results drawn from Table 3, Columns 1, 3 and 5. The Utility term is constructed from a 3-mode Box-Cox Logit model specified in RATE format with distinct transformations applied to its four network variables (Price, Speed, Distance, Frequency) but not to its Socio-economic variables.

14 Results are drawn from Appendices 2, 3 and 4 of the source paper except for the autocorrelation tests which are drawn from Table 11. Note that the Utility term is constructed from 3-mode Box-Cox Logit models by trip purpose (in Table 8 or Appendix 1) each specified with a single transformation common to Fare and In-vehicle time network variables, except in the Vacation trip model that only has a Fare term. The LEVEL 1.4 algorithm (Liem et al., 1993) is used for estimation except in the presence of spatial autocorrelation when an earlier version of the LEVEL 2.1 algorithm (Tran and Gaudry, 2008) is used. The log form of the sum (i.e. the log-sum) is always rejected, as in Part I of this table, even when the actual value of the transformation applied to the Utility term is close to 0 numerically.
Form and difference in forecasts. A simple and global way to compare two versions of a model differing in form is to compare the elasticities evaluated at the sample means, as is done in Table 6 and Table 7. However, it is possible to be more systematic in the comparisons of two forecasts. Consider two variants 1 and 2 of a model differing by assumption only in the functional form applied to each model and assume that we want to see how the forecasts differ with respect to the transport condition variable $X_{qn}$ present in both variants. The two models are, with indices 1 and 2 denoting the models and $n$ an observation subscript:

$$Y_{1n} = \beta_{10} + \sum_k \beta_{1k} X_{kn}^{(\lambda_{1k})} + u_{1n}$$

and

$$Y_{2n} = \beta_{20} + \sum_k \beta_{2k} X_{kn}^{(\lambda_{2k})} + u_{2n}$$

and their difference of interest, neglecting residuals,

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

reaches a turning point with respect to changes in $X_{qn}$ when $\partial \Delta Y_n / \partial X_{qn} = 0$, namely in :

$$\frac{\partial \Delta Y}{\partial X_{qn}} = \beta_{2g} X_{qn}^{\lambda_{2g} - 1} - \beta_{1g} X_{qn}^{\lambda_{1g} - 1} = 0,$$

where

$$X_{**} = \left( \frac{\beta_{1q}}{\beta_{2q}} \right)^{\frac{1}{\lambda_{1q} - \lambda_{2q}}}, \quad \lambda_{1q} \neq \lambda_{2q} \quad \text{et} \quad \beta_{1q} / \beta_{2q} > 0.$$

This means that an analytical value can be calculated for the point of maximum difference $\partial \Delta Y_n / \partial X_{qn} = 0$ between the two forecasts. The interested reader may consult Appendix 9 of Gaudry et al. (2007) where it is shown that, although the crossing point $\partial \Delta Y_n / \partial X_{qn} = 0$ cannot be found analytically but only by solving an equation numerically, $\partial^2 \Delta Y_n / \partial X_{qn}^2 = 0$ the inflexion point of the difference (18) can also be obtained analytically by the expression:

$$\tilde{X}_{qn} = \left[ \frac{\beta_{1q}(\lambda_{1q} - 1)}{\beta_{2q}(\lambda_{2q} - 1)} \right]^{\frac{1}{\lambda_{2q} - \lambda_{1q}}}, \quad \lambda_{1q} \neq \lambda_{2q} \quad \text{et} \quad \beta_{1q} (\lambda_{1q} - 1) / \beta_{2q}(\lambda_{2q} - 1) > 0.$$

B. Away from IIA consistency: the issue of spatial competition

More than one, but how many? But, if endogenous forms give better estimates of the impact of transport conditions on demand, they have no effect on the issue of interdependence among flows: how should it be introduced and tested? IIA consistency is not expected in economics: it does not “take the theory seriously” because theory makes all prices relevant, not just own prices on the diagonal. A difficulty here is that there are $N^2$ flows, or even $(N^2 - N)/2$ if the diagonal is neglected and symmetry imposed. Another is that theory further allows for both substitutes and complements: intervening opportunities models were a way of dealing with complements along paths. We want to propose a parsimonious approach that does not require the inclusion of all prices, but only for the relevant ones, and allows for possible complements among flows.

All-in but excluding complements. So we exclude, as already stated above, the artificial linkage among flows implied by imposed double “additivity” constraints on emissions and attractions, a practice often well documented (e.g. Batten and 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. Taking theory seriously requires Occam’s razor, because not all competing OD pairs matter to explain the flow for a particular one, and the possibility of complements, as in classical demand systems.

**Additivity conditions in another guise, and worse.** But let us count the ways to be profligate, practice indiscriminate “all-in” specifications and to exclude complements at the same time. A first 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 is an economic version (based on expenditures) of the traffic flow additivity condition just noted, but imposes its own heroic assumptions about the distribution in space of expenditure shares that never hold in 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 balances of payments constraints, it is natural to find in flow matrices that do not respect such accounting constraints: (i) when testing for instance with Box-Cox transformations, 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 pound foolishness, such studies simultaneously practice theoretical penny wisdom by imposing substitution between the own flow of interest and all others.

**Parsimony without imposing substitutes: correlation among spatial residuals.** 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, of say (8) or (15), allowing the particular error \(v_i\) associated with an observation \(t\) to be correlated with many values of the same vector \(v_n\), for instance for each of 2 orders:

\[
v_t = \sum_{l=1}^{2} \rho_l \sum_{n=1}^{N} \tilde{R}_{l,n} v_n + w_t
\]

(22)

where \(\tilde{R}_{l,n}\) is the typical element of the matrix \(\tilde{R}_{l}\), a notation which may be clearer in matrix form:

\[
v = \sum_{l=1}^{2} \rho_l \tilde{R}_l v + w ,
\]

(23)

and \(\tilde{R}_l\) 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 (23), 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 \((\rho_1 > 0)\) but that sub-areas within regions could be simultaneously complementary \((\rho_2 < 0)\). As only some O-D 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 (22) into (16).

**From diagonal dominance towards a weighted and distributed form of all-in.** But a 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 may redefine the \(\tilde{R}_l\) matrix as:

\[
\tilde{R}_l = \pi_l \left[ I - (1 - \pi_l) \tilde{R}_l \right]^{-1} \tilde{R}_l, \quad (0 < \pi_l \leq 1)
\]

(24)
where the proximity parameter $\pi_i$ allows endogenisation of the relative importance of near and distant effects considered in the rule of construction of $R_l$. If $\pi_i$ equals one, $\tilde{R}_l$ is equal to $R_l$, indicating that only the adjacent neighbours have an impact on the correlations among the associated residuals assumed in defining the matrix of neighbours $\tilde{R}_l$; this corresponds exactly to a classical case of spatial correlation. By contrast, as $\pi_i$ tends towards zero, the near effect is reduced to a minimum in favour of the distant effect. Generally, the $\pi_i$ 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 $\tilde{R}_l$ structure. This is far more acceptable that full all-in systems where the balancing constraints treat all O-D flows in the same way.

We do not discuss here the results of Tables 6 and 7 obtained with such “long tail” $\pi_i$, estimated with the algorithm documented in Tran and Gaudry (2008), but solely those obtained only with first neighbours defined in the Boolean matrix $R$. In Table 6, first neighbours are defined for any given flow as all those other flows having origins or destinations within 400 miles and 300km, respectively in Columns 4 and 8. In Table 7, the similar rule is 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). Naturally, the log-likelihood gains are much greater when all trip purposes are considered (Columns 5 and 6) than when only the business trip purpose (Column 18) is. In all cases, flows are substitutes because the estimated sign of $\rho$ say so, not by assumption.

**Parsimony without diagonal slavery depends on reality.** Obtaining the right amount of off-diagonality cannot depend on adding-up constraints applied automatically, but on the nature of each case. Clearly, theory does not warrant having all prices in: only those missing from the regression! Such distinctions could be crucial to the modelling of air tourism. Surely, European tourism flows to any tourist area (say the Mediterranean) depend on the prices to some other zones (say, the Atlantic coast, the Baltic Sea and North Africa), and not on all O-D prices and transport conditions in Europe. Modelling reality is hard work.

### 3.3. The Split-type part determining mode/company/scheduled flight and path

Air transport demand modelling does not depend only on how much travel occurs among origin-destination pairs. It depends critically on mode choice. We therefore proceed to show what difference it makes to air mode demand to use a proper mode choice specification. As the workhorse of mode choice analysis is the Logit model, this has to discuss the specification of that model, which is also the critical model of air path choice and air itinerary modelling. In some sense, the discussion of hub demand is often primarily a discussion of Logit mode choice, if one has an interest in the issue at all.

Contrary to the Gravity model, which has a long history and a reference multiplicative (Logarithmic) form, the classical Logit model has short history and a reference Linear form. To explain say the probability of choice (or the market share) of a mode $m$ ($1, \ldots, m, \ldots, M$), it may be written in multinomial garb as:

$$P(m) = \frac{\exp(V_m)}{\sum_{j=1}^{M} \exp(V_j)},$$

with all $M$ utility functions defined as

$$V_i = \beta_{i0} + \sum_n \beta_{i,n} X_{n,i} + \sum_s \beta_{i,s} X_{s,i} + u_i,$$

where we have changed slightly the usual notation to identify the $X_{n,i}$ (the upper index denotes the mode and the lower ones the utility function index $i$ and the network variable $n$), the network characteristics that belong to a particular mode and therefore vary across alternatives, and it is clear that in (26) the $X_s$ denote socio-economic characteristics of consumers that are common across alternatives.

The issue of the nature of competition among modes, paths and company services is everywhere studied almost exclusively with this classical Linear Logit model and its enrichments. We will see that the
explanation of such choices can differ much depending on form, used here both to evaluate better the influence of transport conditions on choices, as above for the Gravity model, but also as a tool to introduce interdependence among alternatives in order to move from diagonal “slavery” to diagonal “dominance” and allow for the possibility of complements.

A. Within IIA-consistent structures: the issue of form in Logit choice modelling

Logit models and the IIA axiom: emergence between 1961 and 1970. Although the logistic curve was discovered and named by the Belgian mathematician Verhulst (1838, 1845) who used it to describe population growth, as did others independently later in the United States (e.g. Pearl and Reed, 1927) and elsewhere, it emerged in binomial canonical form (BNL) only relatively recently. It was first promoted in the bioassay literature (Berkson, 1944) and then in transportation (Warner, 1962), an area suffering from severe underreporting in essays on the history of the Logit model (e.g. Cramer, 2003) excessively focused on the two waves of multinomial development of 1968-1970 and 1977.

For instance, the seminal paper by Abraham and Coquand (1961) on a road path choice multinomial model linked to utility, where the Logit is applied (as an approximation to a Probit) 10 years before it appears in the United States for the same problem (Dial, 1971), should be used by historians of the Multinomial Logit (MNL) as its inception in transportation. It stands out well ahead of the 1968-1970 “first wave” consisting in McFadden (1968 or 1976) on the choice of road tracé and in Theil (1969), Ellis and Rassam (1970), or Rassam et al. (1970, 1971), on the choice of mode.

The multinomial format brings out a feature hidden in the binomial version: the compatibility with the IIA axiom, which translates as equal cross-elasticities of demand. As new transport services or modes never draw proportionately from incumbents, some developments of the Multinomial Logit, such as nested hierarchies, were effectively proposed in 1977 largely to mitigate this excessive property. Nesting hierarchies, however, have their own problems: the different tree structures are not special cases of a general specification —this means that they cannot easily be compared by statistical tests; moreover, they will not properly directly address the structural defect itself, the separability assumption —this requires the reintroduction of off-diagonal terms. Yet, at the time of the introduction of hierarchies, many researchers were thinking of structural reform, as demonstrated by background theoretical work on the Universal Logit and on the Generalized Box-Cox Logit, both outlined further below in Section B: naturally, the reintroduction of off-diagonal terms, often done in ad hoc fashion in ordinary Logit applications undermines the hierarchies precisely by correcting the specification errors on which they are based. Avoiding the unbearable property of equal cross-elasticities of demand means rejecting the coup d’état from which they derive.

Among regal classical demand equation systems, a puritanical coup d’état. Such a model as the MNL, enriched or not by the flowerings of 1977 to be listed shortly, is indeed a coup d’état, because the Logit is simply an additive system of demand equations used during this high growth period after 1968 primarily in transportation but also elsewhere, for instance to explain portfolio shares (Uhler and Cragg, 1971). The demand system background from which it distinguished itself uses ordinal (not cardinal) utility specifications where utility is neither separable, because all prices intervene in all functions, except in Houthakker’s indirect addilog system (1960) which is effectively a logarithmic case of the Box-Cox Logit, nor constant at the margin (functions are of multiplicative (logarithmic) form).

This background notably included at the time the Linear Expenditure system (Stone, 1954) and the popular Rotterdam system (Barten, 1969), to which the Almost Ideal Demand system (Deaton and Mullbauer, 1980) was later added. This classical tradition, allowing for both complements and substitutes, is still much alive (e.g. Barnett and Seck, 2006). So proposing to measure utility cardinally as precisely as the computer chip will allow, making utility separable and dependent only on own transport conditions (on the diagonal) and assuming linearity of the representative utility functions and constant marginal utility, is iconoclastic.

The year of grace 1977. The year 1977 saw an extraordinary flowering of the MNL as three new streams of work simultaneously extended it by: (i) allowing for nested hierarchies (e.g. Williams, 1977); (ii) treating the

15 For instance, the utility functions of public transport modes will suddenly contain the natural logarithm of car ownership (sic) in a model altogether linear and where modal utilities are assumed to be independent!
coefficients in (26) as random instead of fixed (Johnson, 1977, 1978), an approach later known as the «Mixed» Logit; (iii) applying Box-Cox transformations to the explanatory variables (Gaudry and Wills, 1978), a modification first implemented as the “Standard Box-Cox” Logit and later generalized to include off-diagonal terms under the name “Generalized Box-Cox” Logit, as will be seen below.

As generally applied, the hierarchies of nested linear logits do not change the characteristics of the model for our purposes, essentially because they are still IIA consistent across main branches. Similarly, randomization of regression coefficients poses considerable problems of credibility because the distributions of coefficients are unknown most times —does the distribution of a gender variable follow a particular law? is it related to testosterone levels?— and at best unclear due to an information matrix that does not have a closed form, which implies an undefined efficiency bound (Cirillo, 2005). It is of course a formal way of dealing with a type of market segmentation, long practiced in combinatorial fashion by energetic analysts.

More importantly, Orro et al. (2005) have shown in effect that the recent popularity of the Multinominal Mixed Logit (e.g. McFadden and Train, 2000) may well be due to the fact that the true relationships are not linear and should have their curvature estimated rather than postulated. Figure 1 shows the importance of curvature in explaining the response to changes in transport conditions. These authors advocate, for tests of curvature, rewriting (26) in so called Standard Box-Cox Logit utility terms, namely:

\[ V_i = \beta_{i0} + \sum_n \beta_{in} X_n^{(\lambda_{in})} + \sum_s \beta_{is} X_s^{(\lambda_{is})} \]  

(27)

**Form and response to transport conditions in mode choice.** Allowing for non constant marginal utility means that the famous sigmoid shape of the Logit response curve, shown as the dark line of Figure 1, is not symmetric anymore but can have accelerations and decelerations as illustrated. Conversely, asymmetric logistic response curves imply non linearity of the response to the transport condition \( X_n \) graphed on the abcissa. To the extent of our knowledge, the classical form (26) was rejected, in both freight and passenger applications, in all Box-Cox tests made since their first application (Gaudry and Wills, 1978).

**Figure 1. Response asymmetry and non linearity: Classical Linear-Logit vs Standard Box-Cox-Logit**

**Elasticities, signs and marginal rates of substitution.** Form makes a huge difference to modal choice elasticities, as was demonstrated at the time with Figure 2 and Figure 3, among others, where the contrast

---

16 This point was recently brought to our attention by Lasse Fridstrøm.

17 Figures 2 and 3 are the same as Figure 6 and Figure 7 in Gaudry and Wills (1978). In these figures, the single Box-Cox transformation applied to the three transport condition variables is estimated at -0.19. When variable-specific
between classical linear form (on the extreme right of the abscissa) and optimal form (close to the centre, but different from 0 in this case) values are overwhelming. Clearly, modal share forecasts made with an incorrect form have practical implications, as we shall document further and explore presently.

Figure 2. Fare elasticities in intercity Box-Cox Logit model (Canada, 92 city-pairs, 1972)

Figure 3. Travel time elasticities in intercity Box-Cox Logit model (Canada, 92 city-pairs, 1972)

transformations are applied, estimated values are -0.26 (fare), -0.05 (travel time) and 0.57 (frequency); the log-likelihood increases by 5.69 points with the 2 extra degrees of freedom awarded.
And, as the optimal form is often quite different from the initial linear one, sign inversions that correct for previously unreasonable results (e.g. Fridström and Madslien, 1995), or form corrections that correct for silly value of time results obtained under linearity (Gaudry et al., 1989), are a common experience with the Standard Box-Cox Logit model. In fact, contrary to the Gravity form where the multiplicative form is often a very good approximation of the true form, we have not found a single case where the Classical Linear Logit stood up under Box-Cox tests, no doubt because constant marginal utility is untenable.

Any Linear Logit result untested for form is therefore suspicious, because all known tested models to date turned out to have form parameters distinct from both 0 and 1. This is true for instance in Europe-wide passenger models, as with the VAACLAV model (Schoch, 2000) used in TEN-STAC (2004), or with current freight demand models in Germany (Seltz, 2004) or France (Gaudry et al., 2007). It is true in hierarchical structures applied to study High Speed Rail with Canadian intercity passenger flows (Eckbote and Laferrière, 1994) and in European stated preference long-distance freight mode choice studies (e.g. JLR Conseil and Stratec, 2005).

**Specification of transport conditions: expenditure or rate format?** The first application of the Standard Box-Cox Logit illustrated in Figure 2 and Figure 3 demonstrated its usefulness with the transport condition variables of (27) specified in the “EXPENDITURE” format, i.e. as \( f(Fare, Travel\ Time) \), found in classical studies (e.g. Domencich and McFadden, 1975) based on Origin-Destination pair data. But that format is not straightforward: in classical microeconomic demand systems, prices per unit are used as explanatory variables, with Income scaling choices. In this case, with variables specified in the “RATE” specification, i.e. as \( g(Price \ per \ km, Speed, Distance) \), results will differ from those obtained with the expenditure specification only if the utility function is not exactly logarithmic. If it is, the coefficients of the rate specification can obviously be collected to predict exactly those of the expenditure specification and the two models have identical log likelihoods.

In practice, as the optimal form is generally not more logarithmic than it is linear, there is a real choice between these two (non-nested) specifications. The rate specification gave better results in every case we have seen: its also makes microeconomic sense in that Distance plays in it the role of Income in demand systems: Distance becomes a money and time budget line. Furthermore, unit (money or time) prices are less correlated than their expenditure form. Colinearity and specification gains are therefore other features of the Standard Box-Cox Logit specification which make it possible to choose between the expenditure and rate formats of transport conditions. We forecast that the rate format will generally dominate in an increasing number of applications and that some silly forms of market segmentation based on trip length, fare or income will disappear because non linearity is a distinct dimension from segmentation (Algers and Gaudry, 2004).

**Impact of non linearity on forecasts.** These functions have considerable implications for the behaviour of market shares because the marginal impact of price and service characteristics are not constant any more and, as Figures 2 and 3 make abundantly clear, depend on the optimal form. We can therefore go beyond obviously different elasticities or marginal rates of substitution (such as values of time) and beyond other specification and estimation gains to ask, as we did above for Generation-Distribution models, whether anything systematic can be said about the difference between the forecasts from two variants 1 and 2 of a model differing by assumption only in the functional form applied to each model. With indices 1 and 2 denoting these variants and \( n \) an observation subscript, we focus on the transport condition variable \( X_{iun} \) present in the following multinomial \( (i, m = 1, ..., M \) alternatives) variants:

\[
P_{1in} = \frac{\exp V_{1in}}{\sum_m \exp V_{1mn}}
\]

(28)

and

\[
P_{2in} = \frac{\exp V_{2in}}{\sum_m \exp V_{2mn}}
\]

(29)

with representative utility components:
\[ V_{1n} = \beta_{10} + \sum_k \beta_{1k} X_{1kn}^{(\lambda_{1k})} \]  
(30)

and

\[ V_{2n} = \beta_{20} + \sum_k \beta_{2k} X_{2kn}^{(\lambda_{2k})} \]  
(31)

where \( k = 1, K \) denote the independent variables, \((\beta_{10}, \beta_{1k}, \lambda_{1k})\) and \((\beta_{20}, \beta_{2k}, \lambda_{2k})\) are the parameters associated to variants 1 and 2, respectively.

Focusing on the difference between forecasted shares of the \( i^{th} \) alternative \( \Delta p_{in} \) due to changing a variable \( X_{1qn} = X_{2qn} = X_{iqn} \), the so-called own effect [as we neglect here the cross-effect \( \Delta p_{jn} (j \neq i) \)] is given by:

\[
\Delta p_{in} = p_{2in} - p_{1in} = \frac{\exp V_{2in}^{(\lambda_{2k})}}{\sum_m \exp V_{2mn}^{(\lambda_{2k})}} - \frac{\exp V_{1in}^{(\lambda_{1k})}}{\sum_m \exp V_{1mn}^{(\lambda_{1k})}}
\]  
(32)

In contrast with the case previously discussed of Generation-Distribution models where two out of the three points of interest in the behaviour of this difference could be obtained analytically, here\(^{18}\) the **crossing point** \( \Delta p_{in} = 0 \), the **point of maximum difference** \( \partial \Delta p_{in} / \partial X_{iqn} = 0 \) and the **inflexion point** \( \partial^2 \Delta p_{in} / \partial X_{iqn}^2 = 0 \) can only be obtained by simulation, unfortunately [and the same holds naturally for cross effects].

**Examples of the behaviour of \( \Delta p_{in} \) in passenger and freight models.** We will see that simulations have indicated that non linear forms **imply higher market share changes in relatively long trips**, and lower market share changes in short trips, than those forecasted by linear models. This important finding can be demonstrated by calculating, for all observations, \( \Delta p_{in} \) for any mode in a model specified with generic\(^{19}\) coefficients applied to transport conditions: here we use changes of rail transport speed and price based on results found in Table 8 to perform the simulations.

This table presents modal share elasticities and relevant statistics for the two models for which simulations have been made to date. The *ceteris paribus* procedure modifies only one variable in (30) and (31): in the passenger case, train speeds were increased to forecast the effect of the introduction of ICE trains in Germany; in the freight case *intermodal train prices were decreased everywhere by 10%* to determine the effect on all trans-Pyrenean flows.

**Distinguishing between summary and detailed sample-based forecasts.** Starting with the discrete choice passenger model for Germany, we first note in Columns 1 and 2 that the air **time elasticity** evaluated at sample means is 2.16 \([-1.62/-0.75\] times larger at the optimum point \((\lambda = 0.24)\) than at the non optimal point of linearity\(^{20}\) \((\lambda = 1.00)\). But this global statistic hides a structure revealed in Figure 4 where differences between the train share forecasts following the introduction of the faster trains depends on O-D distance or trip length and exhibit an S shape form with a crossing point around 150 km (and hypothetical second one outside of range), an inflexion point around 350 km, and a maximum difference at about 650 km.

---

\(^{18}\) As the interested reader may verify in Appendix 9 of Gaudry *et al.* (2007).

\(^{19}\) If the coefficients of transport conditions were not generic but specific, the exact shape of the S structure of the differences would depend on the mode considered.

\(^{20}\) In Figure 3 pertaining to the aggregate model for Canada, the corresponding ratio is 5.05 \([-0.86/-0.17\] with an optimal form point \((\lambda = -0.23)\) at an equal distance of 0, but negative. Had we considered the comparable ratios for air costs, the ratio in Table 8 is 1.84 \([-0.24/-0.13\] for Germany and 1.94 \([-2.84/-1.46\] for Canada in Figure 2.
Table 8. Selected own share elasticities of Box-Cox Logit models used in the simulations

<table>
<thead>
<tr>
<th>Share / probability elasticity evaluated at sample means</th>
<th>Intercity passengers, Germany 1979, all trip purposes (6000 observations)</th>
<th>Trans-Pyrenean freight flows between the Iberian peninsula and 14 European countries 1999, all freight categories (749 observations)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Column Case Original reference</strong></td>
<td><strong>Mandel et al. (1997)</strong></td>
<td><strong>1 Linear MKII</strong></td>
</tr>
<tr>
<td>Fare (per O-D)</td>
<td><strong>Price (per ton-km)</strong></td>
<td><strong>Intermodal train</strong></td>
</tr>
<tr>
<td>Elasticity</td>
<td>Plane</td>
<td>-0.99</td>
</tr>
<tr>
<td></td>
<td>Train</td>
<td>-0.13</td>
</tr>
<tr>
<td></td>
<td>Car</td>
<td>-0.04</td>
</tr>
<tr>
<td>t-statistic of $\beta_k$</td>
<td>(-8.39)</td>
<td>(-5.48)</td>
</tr>
<tr>
<td>Speed (km/h)</td>
<td>Plane</td>
<td>-0.75</td>
</tr>
<tr>
<td></td>
<td>Train</td>
<td>-0.63</td>
</tr>
<tr>
<td></td>
<td>Car</td>
<td>-0.08</td>
</tr>
<tr>
<td>t-statistic of $\beta_k$</td>
<td>(-14.92)</td>
<td>(-15.14)</td>
</tr>
<tr>
<td>Other variables</td>
<td>(not reported on here)</td>
<td>(not reported on here)</td>
</tr>
<tr>
<td>Box-Cox $\lambda$ on Fare and Time</td>
<td>1.00</td>
<td>0.24</td>
</tr>
<tr>
<td>Box-Cox $\lambda$ on Price</td>
<td>Log-likelihood</td>
<td>-1230</td>
</tr>
<tr>
<td>Box-Cox $\lambda$ on Distance</td>
<td>Degrees of freedom</td>
<td>0</td>
</tr>
</tbody>
</table>

**Used in simulations shown in Figures 4 (ICE added) and 5 (container trains)**

Considering now in Table 8 results of the aggregate model of European trans-Pyrenean freight flows, we note in Columns 3 and 5 that the rail **price elasticity** evaluated at the sample means is 6.30 [= -6.30/-1.34] times larger at the point of optimal form ($\lambda = -1.83$) than at the non optimal point of linearity ($\lambda = 1.00$). And we see in Figure 5 that, when intermodal container train prices are reduced by 10% on all origin-destination pairs, these overall statistics also hide differences among forecasts related to trip length in S-shaped manner, with crossing points around 600 and 2500 km, an inflexion point around 1400 km, and a maximum difference at about 1700 km. As compared to Figure 4, there are two crossing points here because, given parameter values, the range of distances pertaining to trans-Pyrenean freight flows linking origins and destinations within Europe (we excluded the small numbers of flows from Morocco to Poland) is sufficiently large to allow it, by contrast with the case of domestic German passenger flows.

**Revenues depend on the product of Generation-Distribution and Mode Choice models.** Such findings have major implications for revenue generation because short trips are more numerous than long trips, as determined in Generation-Distribution models above where elasticities with respect to distance are even typically greater than one [in absolute value], implying a faster than proportionate decrease. This will be true whether modes or itineraries are considered: non linear forms seem to increase the market shares effects of changes in transport conditions for relatively long trips and decrease those of relatively short length.

Eckbote and Laferrière (1994) examined such revenue implications with a nested Logit specification applied to stated preference survey data pertaining to High Speed Rail (HSR) options in the Quebec-Windsor corridor of Canada and found, after setting revenue maximizing prices for each of the two trip purposes, that the dominant non linear forms [with optimal $\lambda$ values found for price at 0.25 (business) and 0.31 (other purposes) and also duly estimated separately for travel time and access time (but not for frequency of service)], implied lower HSR revenues with optimal forms than those obtained with the (non optimal) linear form of the same models. We now turn to another study of the same corridor as we try to introduce off-diagonal terms in the utility functions to go beyond the use of own characteristics in diagonal structures.
B. Away from IIA consistency in mode choice: diagonal dominance instead of diagonal slavery

Theory: early but lonely. We indeed wish to show, by outlining alternatives to diagonal slavery and consequent IIA, that it is possible to bring the Logit back into the fold of demand systems and to take interdependence of utility into account by using Box-Cox transformations on the additional terms that, if theory is “taken seriously”, should differ from those used on the diagonal: were the optimal forms the same, the enriched model would not be identifiable. As we will see in both Table 9 and Table 10 below, identification does not pose undue problems as the estimated values of Box-Cox transformations shown in bold indicate.

We outlined this strategy in detail early enough (Gaudry, 1978) after the flowering of 1977 had, in our view, shown the unsustainability of linear Logit utility functions, the practical manageability of Box-Cox forms
applied to diagonal terms\textsuperscript{21} and to socio-economic terms: their possible use to obtain a workable specification of the Universal Logit, originally specified at a high level of generality by McFadden (1975), came naturally.

But adding the characteristics of competing alternatives in own mode utility function did not appeal to everyone: some researchers (e.g. Ben-Akiva, 1974) feared to find unexpected results, like complementary modes, or colinearity as in the classical systems. To proceed carefully, we start in Table 9 with a partial (passenger) model, where only the car travel time variable is used in all utility functions, and then proceed to further develop a full price matrix model, on the basis of the freight example just reported on.

\textbf{From diagonal slavery to a weak form of diagonal dominance.} The results shown in Table 9 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 dark 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\textsuperscript{22} (from Column 1 to 2 and from Column 4 to 5) and the incorrect sign is corrected\textsuperscript{23}. Clearly, it is useless and incorrect to avoid the problem by constructing a generalized cost variable for non-business trips, as was done by those who first analyzed this high quality Via Rail Canada database of 12 000 trips with (nested or not) linear Logit forms (KPGM, 1990), or to ignore this trip purpose entirely (Bhat, 1995); (iii) 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. This completely addition modifies the error correlation structures which arbitrary nests are meant to address and in fact provides a continuous alternative to non-nested nesting structures; it also certainly modifies the variance of the error terms that may or may not be originally homoskedastic\textsuperscript{24} in linear space, an issue beyond our concern here but discussed elsewhere jointly with that of the estimation of Box-Cox transformations (Gaudry and Dagenais, 1979).

Also, the values of time obtained in Table 9 (measured in 1987 Canadian dollars), differentiated by mode, make more sense than the unique ones of the linear form. It should also be noted that all off-diagonal term elasticities are strong and their t-statistic results (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”. We also note in passing the presence on non linearity with respect to Income.

Although our emphasis is on off-diagonal transport condition terms, it should be mentioned that Box-Cox transformations make it possible to solve also in principle the under-identification problem associated with coefficients of socio-economic variables, a problem shared with alternative-specific constants [discussed below with equation (34)], in linear formats. In Table 9, alternative-specific effects are estimated (with the car mode as reference) with Income successfully transformed, in contrast with the situation found in Table 10 where generic regression coefficients (with a truck reference) are estimated for the Distance variable (also successfully transformed) and for the Market size variable (linear after tests). In no case presented here are alternative specific coefficients estimated for all of the alternatives, although this was tried without success for the Market size variable and is clearly of interest.

\textsuperscript{21} The estimation algorithm was written and distributed early (Liem and Gaudry, 1987, 1993).

\textsuperscript{22} It is possible to show that the only transformation which is not significantly different from 1, the linear case, is that of the Frequency variable.

\textsuperscript{23} Clearly, as the results of Table 9 demonstrate again after Fridström and Madslien (1995), sign change is a critical issue linked to how covariances among variables depend on their form. This means that models of untested form are suspicious and that it is inadequate to assume linearity “for simplicity”, as is still regularly done (e.g. Berri et al., 2004).

\textsuperscript{24} Heteroskedasticity, as it should be properly spelled in English (McCullogh, 1985), was studied by Bhat (1995) using only the business trip purpose data and dropping entirely the bus mode (3264 observations), an entirely arbitrary stance to avoid facing the sign issue with non business trips, and perhaps problems of heteroskedasticity of the bus utility.
We did not have time to systematically introduce car cost and other modal characteristics, and did not wish to do so, because the EXPENDITURE format of the utility functions that we inherited from the first analysts were never compared by them to a RATE format. This is normally a second step performed only after it has been demonstrated that the optimal forms are not logarithmic, but one could envisage a comparison of (non nested) linear forms of the EXPENDITURE and RATE specifications. In the next example discussed

\[ \text{Table 9. Comparing Linear with Standard and Generalized Box-Cox elasticities and values of time} \]

<table>
<thead>
<tr>
<th>Quebec City-Windsor corridor of Canada 1987 (domestic intercity flows)</th>
<th>Business trips (4 402 observations)</th>
<th>Other trips (8 535 observations)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighted aggregate probability point elasticity</td>
<td>Column Case</td>
<td>1 Linear</td>
</tr>
<tr>
<td>Original reference</td>
<td>Model 3</td>
<td>Model 5</td>
</tr>
<tr>
<td><strong>Cost</strong> (access + in-vehicle)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Plane own cost</td>
<td>-0.53</td>
<td>-0.46</td>
</tr>
<tr>
<td>- Train own cost</td>
<td>-0.05</td>
<td>-0.13</td>
</tr>
<tr>
<td>- Bus own cost</td>
<td>-0.03</td>
<td>-0.13</td>
</tr>
<tr>
<td>- Car own cost</td>
<td>-0.05</td>
<td>-0.10</td>
</tr>
<tr>
<td>( t )-statistic of ( \beta_k )</td>
<td>(-3,66)</td>
<td>(-7,23)</td>
</tr>
<tr>
<td>Associated Box-Cox ( \lambda_k )</td>
<td>1.00</td>
<td>0.28</td>
</tr>
<tr>
<td><strong>Travel time</strong> (access + in-vehicle)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Plane own time</td>
<td>-0.32</td>
<td>-0.11</td>
</tr>
<tr>
<td>- Train own time</td>
<td>-0.18</td>
<td>-0.10</td>
</tr>
<tr>
<td>- Bus own time</td>
<td>-0.26</td>
<td>-0.14</td>
</tr>
<tr>
<td>- Car own time</td>
<td>-0.11</td>
<td>-0.05</td>
</tr>
<tr>
<td>( t )-statistic of own ( \beta_k )</td>
<td>(-10,19)</td>
<td>(-5,91)</td>
</tr>
<tr>
<td>( t )-statistic of cross ( \beta_k )</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Associated own generic Box-Cox ( \lambda_k )</td>
<td>1.00</td>
<td>1.80</td>
</tr>
<tr>
<td>Associated cross generic Box-Cox ( \lambda_k )</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td><strong>Frequency</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Plane own frequency</td>
<td>0.39</td>
<td>0.46</td>
</tr>
<tr>
<td>- Train own frequency</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>- Bus own frequency</td>
<td>0.12</td>
<td>0.10</td>
</tr>
<tr>
<td>( t )-statistic of specific ( \beta_k )</td>
<td>(12,52)</td>
<td>(11,67)</td>
</tr>
<tr>
<td>Associated generic Box-Cox ( \lambda_k )</td>
<td>1.00</td>
<td>1.53</td>
</tr>
<tr>
<td><strong>Income</strong> (Gross Individual)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Plane (reference: car)</td>
<td>0.23</td>
<td>0.18</td>
</tr>
<tr>
<td>- Train (reference: car)</td>
<td>-0.04</td>
<td>-0.04</td>
</tr>
<tr>
<td>- Bus (reference: car)</td>
<td>-0.10</td>
<td>-0.10</td>
</tr>
<tr>
<td>( t )-statistic of specific ( \beta_k )</td>
<td>(-2,61)</td>
<td>(-2,99)</td>
</tr>
<tr>
<td>Associated generic Box-Cox ( \lambda_k )</td>
<td>1.00</td>
<td>0.16</td>
</tr>
<tr>
<td><strong>Other variables not reported</strong></td>
<td>(Trip origin in a large city)</td>
<td>(Trip origin in a large city; party size)</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-1068</td>
<td>-1058</td>
</tr>
<tr>
<td>Degrees of freedom</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

We did not have time to systematically introduce car cost and other modal characteristics, and did not wish to do so, because the EXPENDITURE format of the utility functions that we inherited from the first analysts were never compared by them to a RATE format. This is normally a second step performed only after it has been demonstrated that the optimal forms are not logarithmic, but one could envisage a comparison of (non nested) linear forms of the EXPENDITURE and RATE specifications. In the next example discussed

---

25 The probability point elasticity is the usual elasticity of the probability multiplied by the probability of the alternative. For details, see Liem and Gaudry (1987, 1993).
presently, the correct format was determined (the RAte format won easily) and extensive tests were carried out to find the best sequence to introduce off-diagonal terms: most significant additions on a one-to-one basis constitute the “first round” of interdependence terms, and so on for less significant additions, with the resulting sequence defined already in Table 5 above.

Table 10. Linear, Standard and Generalized Logit mode choice model of freight across the Pyrenees

<table>
<thead>
<tr>
<th>Flows between the Iberian peninsula and 14 European countries 1999, all freight categories, by land across the Pyrénées (749 observations)</th>
<th>Progressive inclusion of Price variables (added to Speed, Distance and Total O-D market size)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model numbers 58 to 62 are as in the source: Gaudry et al. (2007)</td>
<td>Road</td>
</tr>
<tr>
<td>Constraints from Table 5 and λ on Distance</td>
<td>V_i utility function</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>D. F.</td>
</tr>
</tbody>
</table>

A. Linear Logit Model (58)

| -2581 | 0 | 1.00 | 1.00 | 1.00 | 1.00 | V_road r | -0.382 |
| V_container c | -0.194 |
| V_classic rail f | -0.137 |

B. Standard Box-Cox Logit Model with one BCT (59)

| -2550 | 1 | -1.75 | 1.00 | 1.00 | 1.00 | V_road r | -0.517 |
| V_container c | -0.933 |
| V_classic rail f | -0.093 |

C. Standard Box-Cox Logit Model with two BCT (60)

| -2535 | 2 | -1.83 | 1.00 | 1.00 | 6.49 | V_road r | -0.523 |
| V_container c | -0.984 |
| V_classic rail f | -0.093 |

D. Generalized Standard Box-Cox Logit Model with first round of substitutes or complements (61)

| -2524 | 4 | -1.709 | -6.271 | 1.00 | 6.272 | V_road r | -0.572 |
| V_container c | 0.441 |
| V_classic rail f | 0.079 |

E. Generalized Standard Box-Cox Logit Model with all rounds of substitutes or complements (62)

| -2510 | 6 | -1.89 | 1.05 | 0.16 | 6.92 | V_road r | -0.489 |
| V_container c | 0.383 |
| V_classic rail f | 0.068 |

A stronger form of diagonal dominance: a practical form of the Universal Logit. Although the model considered has Speed, Distance and a measure of Total O-D market size in the utility functions, our emphasis here is on prices, as documented in Table 10. To be formal, the inclusion in the utility functions of the characteristics of other alternatives, in the so called Generalized Box-Cox Logit form, modifies (27) to:

\[ V_i = \beta_0 + \sum_n \beta_{in} \cdot X_i^{(\lambda_{in})} + \sum_n \beta_{jn} \cdot X_j^{(\lambda_{jn})} + \sum_s \beta_{js} \cdot X_s^{(\lambda_{js})} \]  

(33)

which permits complex pattern of substitution and complementarity, as classical demand systems also allow. We must therefore comment on the evolution of the sizes and signs of price own and cross-elasticities of the market shares.

We have already noted above (in our comments of Table 8) that passing from (A) to (B) by putting a Box-Cox transformation λ_i on all own prices on the diagonal of the price matrix has an important effect on own share elasticities, as reiterated in Table 10 where, by contrast, almost no impact on the Price elasticities is found in (C) after the addition of the second Box-Cox transformation λ_i to the Distance term, despite the fact that this second transformation itself increases the log likelihood by 15 units and obtains a very strong value.

**Note that, although the first three models A, B and C are the same as those reported on in columns 3, 4 and 5 of Table 8, the elasticities found here differ: in Table 8 all elasticities presented are the normal share elasticities. The latter, multiplied by the market shares, yield the percentage point elasticities, as explained in the SHARE program documentation (Liem et al., 1997) where the interested reader may verify that all derivatives of shares with respect to any variable duly take into account its presence in all utility functions where it appears.

Tests made on modal Speed showed that its effect was approximately linear.
Can transport modes be gross complements? Concerning specifically the introduction of interdependence in utility, note further that the addition in (D) of the first round of three off-diagonal terms naturally yields massive gains (they were tested one at the time on the basis of their individual effect on the log likelihood value) but provides no indication of complementarity because all off-diagonal terms are positive. The off-diagonal term effects are, in the majority of cases, much smaller that those of the diagonal terms; in the case of the classic rail utility function where they are about equal, the t-statistic of the road price coefficient underlying 0.13 is much lower than -11.83, that of the own price underlying -0.12. Overall, this suggests that, as in demand systems, off-diagonal substitution effects are weaker than diagonal effects which “dominate” them, as they are expected to do in demand systems: it would be a weird case if off-diagonal term cross-elasticities were stronger than diagonal term own elasticities in a model where many such “first round” terms intervene.

How many off-diagonal terms? Can one go further and add a second round of prices? When the price matrix is full in (E): (i) the elasticity values on the diagonal hardly budge, as compared to those in (D); (ii) all off-diagonal terms are positive except for -0.145 [with a t-statistic of (9.55)] of the rail price in the intermodal utility function, which is negative like the own elasticity of -1.233 [with a t-statistic of (-8.18)]. One might then be led to think that, because intermodal rail combines road and rail components, this complementarity (of classic rail to intermodal rail) result is reasonable despite the fact that the converse is apparently not the case —intermodal rail is not a complement of classic rail because its reverse elasticity is positive and equal to 0.068 [with a t-statistic of (-0.037)]. Of course, t-statistics have to be used to weigh all results: the converse complementarity effect is not significant.

These results therefore imply the presence of some complementarity but the (fragile) asymmetry of cross-effects between container and classic rail poses some questions. Another suspicious result is that the elasticity of road demand with respect to the price of combined intermodal transport is higher in absolute terms [at 1.165, with a t-statistic of (11.05)] than the own elasticity [at -0.489, with a t-statistic of (8.18)]. Are we at the limit of the technique and encountering the same difficulties as classical demand systems: colinearity? Will constraints like (Slutsky-type) symmetry conditions sadly have to be imposed?

An careful analysis of the best colinearity indices, those of Belsley et al. (1980), indicates a serious deterioration of the indices as one goes from (D) to (E). This means that, despite obvious robustness on the diagonal, some of the cross effects are imprecisely estimated. We therefore consider the (E) results as lacking in robustness and, short of a deeper analysis, will not use them to answer the next question.

Impact of additional non linear off-diagonal terms on forecasts. What does the presence of cross terms imply for systematic differences between linear and non linear forecasts? If the diagonal effects dominate, one should not expect the additional off-diagonal terms in (D) to modify what was found earlier in Figure 5 as we proceeded from linear to Standard Box-Cox forms. Note also that the difference \( \Delta p_{in} \) between the two forecasts is still expressed by (32). But, as the variable \( X_{qn} \) now appears in all alternatives, its modal index \( i \) is not necessary anymore and the own and cross elasticity formulas are per force identical. It is also true here that the crossing point \( \Delta p_{in} = 0 \), the point of maximum difference \( \partial \Delta p_{in} / \partial X_{qn} = 0 \) and the inflexion point \( \partial^2 \Delta p_{in} / \partial X_{qn} = 0 \) can only be obtained by simulation, as held in the Standard case. Figure 6 exhibits the same behavioural pattern as Figure 5 despite the fact that model (61) replaces model (60).

28 The t-statistics pertaining to Table 10 are found in Table 7 and Appendix 3 or 6 of the source document.
29 This Generalized Box-Cox Logit is the model in use since 2005 by the French ministry of transport to forecast freight land mode market shares across the Pyrénées. This ministry now frequently requires Box-Cox tests in call for tenders pertaining to demand models.
The presence of reasonably weak off-diagonal effects does not change the structure of results already obtained with the Standard Box-Cox formulation. Non linearity therefore make the strongest difference to our problem—implying a sigmoid behaviour in the difference between the linear and non linear forecasts—despite specification enrichment: we are back in classical demand systems where all prices matter and utility is not presumed to be separable, but the diagonal dominates without enslaving us, and the full matrix contains a trace of modal complementarity—another entirely new result attributable to an excellent database.

C. Implications for path, company or service choice in air markets

**Company, schedule and path choice.** As a cross-sectional problem, the choice of company, scheduled flight and passenger itinerary is not different from the mode choice problem just discussed at length. Generally, path choice models are of the Linear Logit kind, with the same properties as in mode choice applications except one: that of the constants, to be discussed shortly. Until as recently as 2004, Lufthansa was using a Linear Logit procedure for path choice problems. As the number of paths used by air passengers tend to increase with distance (from say 3 within Germany up to 16 between Germany and the United States), the modelling procedure is of some import.

The emergence of the non linear Box-Cox Logit model implies that some paths will incorrectly dominate other paths under a linear form, but these effects have yet to be studied in comparative fashion as they are expected to mimic those found for the mode choice models noted above. They have implications for the appropriate size of planes: linear forms would put larger planes than necessary on relatively short flights and planes of somewhat insufficient size on relatively long flights, as the horizontal S-shaped curve of differences in forecasts between models implies in Figures 5 or 6.

**Efficient Stated Preference sample design.** Use of non linearity of the Box-Cox type should therefore have a strong impact on marketing studies and Stated Preference surveys: it implies that experimental data should contain sets of variables that are orthogonal not in linear but in non linear space. Indeed, the correlation that matters for the reproduction of reality and the establishment of statistical causality is defined among the optimal nonlinear forms of the variables of interest, a point particularly well made in Cirillo (2005).

In the analysis of air carrier and path choice, there are currently few exceptions to the use of a Linear Logit and such exceptions as exist, such as the PODS procedure derived from an earlier Boeing (Boeing Co, 1996) procedure, works very much like a Linear Logit (Carrier, 2003), no doubt partly because its “diagonal”
utility functions are linear, as they are in complex variants of the Multinomial Logit applied to air path choice (Walker & Parker, 2006; Djuricic, 2007). Importantly, PODS also attempts to take into account capacity constraints on links, which artificially generates some complementarity among them, in the spirit of the seminal MAPUM (Soumis, 1978; Soumis et al., 1979) air path choice algorithm.

The many ways to avoid interdependence and develop compensating mechanisms. This “linear logit-like + overflow and overspill” practice should bias decisions towards high frequency hubs and stabilise path assignments by motivating the development of complementary block groupings (such as Toronto-Paris, Montreal-Paris and Montreal-Toronto-Paris) in part because of the simplistic linearity assumptions preventing reference paths from affecting the utility of all alternatives. In many of these problems, this “reference” choice (shortest path at desired departure time, etc.) is like the car time characteristic of the Quebec-Windsor mode choice model developed to determine the market for high speed rail in central Canada: it belongs in all utility functions because the utility of any alternative is not separable from that of the reference alternative.

The problem of constants in unimodal path assignments, notably air networks. In this context, a problem with the Logit is that its mode-specific constants are underidentified: consequently one ends up estimating differences in their coefficients with respect to an arbitrarily chosen reference one, as in:

\[
V_i = (\beta_{i0} - \beta_{r0}) + \sum_n \beta_{in} X_n^i + \sum_s (\beta_{is} - \beta_{rs}) X_s
\]

\[
V_i = \beta_{0i}^v + \sum_n \beta_{in}^v X_n^i + \sum_s \beta_{is}^v X_s
\]  

(34)

where the problem is illustrated for these constants and for the socio-economic variables common to all functions.

This problem is minor in mode choice problems if no new mode is to be introduced, but in company/path choice procedures, it matters whether the path constants are equal because the game is precisely to introduce new paths all of the time. The reason why this is a hard problem in air networks is that there is no natural labelling of paths (first, second, third, etc.) serving an O-D pair, as there is in mode choice analysis (bus, plane, car, train, etc.), because it hardly makes sense to define a reference path as one might define a reference mode: it has to come out of the analysis. Some research is going on in this area. The issue is finessed by some path analysts (Abraham and Coquand, 1961; Dial, 1971) by setting all constants at 0 or by ignoring it while working on the choice of road tracé (McFadden, 1968).

Constants and ignorance parameters. One way out of this constant identification quandary is to use an Inverse Power Transformation (IPT) envelope (Gaudry, 1981) applied to a Logit (IPT-L) path choice model, as was done by Laferrière (1987) to identify all his path-specific air constants for a Canada-wide model built with a 100% sample (16 million individual trips) of domestic air trips made on Air Canada and Canadian Pacific Airlines in 1983. If one further accepts a defined ordering of air paths as a testable device, reference alternatives should be identifiable.

Moreover, some additional gains can be expected from the IPT-L form, such as the identification of captivity to certain paths (or airlines and airports) and the detection of leftover non linearity or asymmetry in the response to such ordered path utilities. Captivity, or modeller’s ignorance, amounts to non 0 or 1 asymptotes of the response curves in Figure 1. It is distinct from the issue of its asymmetry. This new orientation is therefore promising for the proper modelling of air company/path competition that generates airport demand.

4. Conclusion: a TAM and remedies to build in Interdependence among Flows

A brand new TAM. In this paper, we first defined a new four-part Traffic Accounting Matrix (TAM) to register all spatial flows of interest for air demand forecasting, effectively extending the scope of classical algebraic input-output analysis by doubling up and reinterpreting the intermediate and final transactions components of two-part Input-Output (I-O) matrices. Strictly defined subsets of a TAM can then be matched to, and explained by, the usual procedures pertaining to the distinct generation, distribution, mode choice and assignment steps of traffic demand planning, or to their combinations.
**Good curvature and proper statistical correlation are insufficient to dump IIA.** We then focused on the key properties of such demand models in order to evaluate their relevance to the explanation of airport or hub competition and considered, among potential remedies, the estimation of form with Box-Cox transformations, but pointed out that their demonstrated relevance to the measurement of the impact of transport conditions is insufficient to solve the problem at hand. In both Generation-Distribution and Split Choice mode-company-path structural steps, the predominant use of Independence from Irrelevant Alternatives (IIA) consistent cores must be rejected to account properly for competition among destinations in Generation-Distribution models and for the prevailing importance of reference alternatives in Split Choice mode-company-path models.

**But form matters for transport conditions.** We provided a first partial literature summary of numerous results obtained with endogenous functional forms in both of these structural steps but argued that, because the issue of non separability of utility is not directly addressed by standard Box-Cox transformations, their increased explanatory power and realism—as compared to the popular fixed form *a priori* logarithmic (in Gravity models) and linear (in Logit models) specifications— is more relevant to the proper measurement of the role of transport conditions (distance, level of service or price) than to the necessary representation of interdependence among alternatives, which mandates the abandonment of “diagonal slavery” in utility formulations.

**Form corrects mistakes and is consistent with non constant marginal utility.** Of course, the proper role of transport conditions still matters decisively in both Generation-Distribution models of transport or trade and in the Mode or company-path Choice splits. In the former class, proper curvature defines the total market reach and data determined forms rectify the demonstrably incorrect use of distance in the many logarithmic pooled time-series and cross-sectional models. In the latter class, allowing for changing marginal utility profoundly modifies the relative sensitivity of longer over shorter length trips, as compared to their behavior in prevailing untested linear constant marginal utility forms of the same functions, never theoretically very credible nor empirically sustainable. But none of these benefits and remedies to current dominant practice allows for interdependence (non separability) of utility, the key future demand modelling challenge if *ex ante* forecasts are to be of relevance to the air demand question at hand.

**But introduction of non diagonal terms is necessary.** To point to real remedies, we summarized some recent promising attempts to deal with interdependence in manageable ways expected to yield “diagonal dominant” results: through the use of spatial autocorrelation in Box-Cox Generation-Distribution models and of Generalized Box-Cox specifications in Split models. Separable utility is thereby rejected by the data but without using too many independent off-diagonal terms pertaining to transport conditions: if the denial of any separability has been the scourge of classical demand equation system, its blind imposition has been that of Gravity and Logit demand systems. Considerate and flexible middle ways are now within reach and they matter most to model new interdependent markets, such as tourism.
5. References


Gaudry, M. and A. Le Leyzour, "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, 1994.


Liem, T. C. and M. Gaudry, “QDF: a Quasi-Direct Format Used to Combine Total and Mode Choice Results to Obtain Modal Elasticities and Diverstion Rates”, Publication CRT-982, Centre de recherche sur les transports, Université de Montréal, 19 p., June 1994.

Liem, T.C., Gaudry, M. and R. Laferrière "SHARE: The S-1 to S-5 programs 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 CRT-899, Centre de recherche sur les transports, Université de Montréal, 70 pages, 1997.

Lill, E., Das Reisegesetz und seine Anwendung auf den Eisenbahnverkehr, Vienna, 1891.


Quesnay, F., Tableau Oeconomic, 1 page, Versailles, 1ère édition, 1758.


6. Annex 1. The simple cost and frequency analytics of hubbing

What happens to route length and flight frequency when one changes from bilateral to hub structures? Imagine a country A with 5 cities, located at its centre and corners, and a second country B of hexagonal form with 7 cities similarly located, as in Figure 7.

![Figure 7. Two spatial configurations](image)

In each case denote by $c$ the length of the link from a corner city to the central city. Now compare the length of a network of direct non-stop bilateral links among all cities to a hub-shaped network centred on the central city. For country A, the hub-based network is 3.41 times shorter than the other and, for country B, it is 4.73 times shorter.

Service frequency will increase correspondingly when, using the same planes, the grid structure is replaced by the hub structure. To see this, first note that, in each case the number of bilateral links among $n$ cities is $n(n-1)/2$. In each case the length of the hub-shaped network is $(n-1)c$ and the corresponding length of the grid-shaped network is equal to $(n-1)c((n-1)/2 + ((n-1)/2)^{1/2})$.

In each country, the obviously positive difference between the two configurations is $(n-1)c ((n-1)/2 + ((n-1)/2)^{1/2} -1)$. To further compare actual gains between cities, one would have to further specify that length $c$ is equal for both cities, but this has not been assumed in this case. One could study this difference further by taking derivatives with respect to $n$, but that is unnecessary here for our purposes.