EFFECTS OF ROAD GEOMETRY AND SURFACE ON SPEED AND SAFETY
A first simultaneous non linear equations analysis distinguishing between risk and uncertainty
by

Marc Gaudry[1,2]
Karine Vernier[3]

1 Agora Jules-Dupuit
ajd@ajd.umontreal.ca
marc.gaudry@umontreal.ca

2 Département de sciences économiques *(department of economics)*
Université de Montréal

3 Gaz de France (GDF)
Paris
vernier@gazdefrance.com

This study has received an initial support from Émile Quinet, a three year financing, and the data bases of the Service d'études techniques des routes et autoroutes *(SETRA - Roads and Motorways Engineering Department)* of the du Ministère de l'équipement, du logement, des transports et du tourisme *(MELTT - Department of the Environment, Transport and the Regions)* of France as well as a merit scholarship of the Government of Canada obtained by Karine Vernier during a year at the Centre de recherche sur les transports *(CRT - Centre for research on transportation)* of the Université de Montréal. Marc Gaudry's contribution has been continuously supported by the Société de l'assurance automobile du Québec *(SAAQ -Quebec automobile insurance society*) within the framework of the international network of DRAG type models and by the Natural Sciences and Engineering Research Council of Canada (NSERC). In 1998, this contribution has also been supported by the position as a researcher at the Centre national de la recherche scientifique *(CNRS - French National Center for Scientific Research)* at BETA, Louis Pasteur University and UMR CNRS 7522, and more specifically by a stay as a foreign researcher invited at the Département évaluation et recherche en accidentologie *(DERA – Accidentology evaluation and research department)* of the Institut national de recherche sur les transports et leur sécurité *(INRETS - The French National Institute for Transport and Safety Research)* of Arcueil. The authors wish to thank Jean-Marie Dufour, Marie-Thérèse Goux, François Laisney, Sylvain Lassarre, Thomas Lemieux and Urs Maag for their precious advises and Francine Dufort, Piotr Tarnovski and Liem Tran for their support linked to the use of the TRIO software used at the BETA, the CRT, the INRETS and the SETRA. The latter is in no way responsible for the use made of the information provided to the authors, who are entirely responsible for the results obtained and wish that the supplementary information eventually available on the variables poorly explained in the current data bases will transform this exploratory model into a working tool.

Agora Jules Dupuit – Publication AJD-18

November 1998. Revised June 2002

Résumé

Le présent article propose une méthode d'analyse des comportements routiers vis-à-vis de la sécurité routière et de la vitesse en fonction du trafic ainsi que des caractéristiques de géométrie et de surface routières. Nous élaborons une structure simultanée permettant de tenir compte des différents ajustements entre confort, vitesse et incertitude liés aux préférences des usagers. Dans l'esprit de Frank Knight, la formulation adoptée distingue, parmi les composantes du risque observé ou objectif, entre l'incertitude ou dangerosité et le risque calculé associé à la vitesse. On y utilise aussi pour la première fois une mesure empirique nouvelle du “ risque perçu ”, l'espérance maximale d'insécurité EMI, dérivée de la théorie de l'utilité aléatoire, avec précisément l'incertitude comme composante centrale, qui permet d'isoler le rôle de ce facteur parmi les autres déterminants du choix de vitesse. La structure d'estimation repose sur une modélisation de la vitesse (moyenne et variance) par un modèle non linéaire et de la sécurité routière (fréquence et gravité) par des modèles de choix discret de type Logit autorisant à la fois non linéarité et stratification endogène de l'échantillon. Il serait donc possible, sans engagement de frais trop importants, de compléter la base de données utilisée pour y inclure les variables désirées et corriger d'éventuels biais présents si ces variables n'étaient pas orthogonales à celles dont nous disposions.

Mots clés: accidents de la route, fréquence, gravité, vitesse moyenne, écart type des vitesses, structure multiniveaux DRAG, simultanéité, Frank Knight, incertitude, risque, espérance maximale d'insécurité EMI, infrastructure, tracé, géométrie, surface, structure de la chaussée, non linéarité, méthode du maximum de vraisemblance , Box-Cox, Logit, stratification endogène d'échantillons, progiciel de régression TRIO, points de probabilité, élasticités, France, stations SIREDO, SETRA, MELTT.

Abstract

This paper presents a method to analyze the user's attitude toward speed and road safety in relation to traffic, road design and surface characteristics. A simultaneous structure makes it possible to take into account the trade-offs among comfort, speed and uncertainty dependent upon user preferences. Our formulation notably distinguishes between two components of observed objective risk, namely calculated risk linked to speed, and uncertainty or “ dangerousness ”, in the manner of Frank Knight. We also use for the first time a new empirical measure of “ perceived risk ”, expected maximum insecurity EMI, derived from random utility theory, with uncertainty at its core, facilitating the identification of the role of perceived risk among the determinants of speed choice. The estimation structure consists of three equation groups : the first two explain accident frequency and severity with discrete choice logit-type models easily admitting of nonlinearity and choice-based sampling. The speed equations, explaining both the mean and the variance of speeds, also consist of non linear flexible-form models. Our choice-based sampling approach permits future completion of the data base at low cost, in order to correct potential biases due to missing variables that may not be orthogonal to those at our disposal.

Keywords: road accidents, frequency, severity, average speed, standard error of speeds, multilevel DRAG structure, simultaneity, Frank Knight, uncertainty, risk, expected maximum insecurity EMI, infrastructure, layout, geometry, surface, roadway structure, non linearity, maximum likelihood method, Box-Cox, Logit, choice-based sampling, TRIO regression package, probability points, elasticities, France, MELTT, SETRA, SIREDO stations.

Zusammenfassung

In diesem Artikel wird ein Verfahren vorgestellt, das es erlaubt, die Einstellung des Nutzers zu Fahrgeschwindigkeit und Verkehrssicherheit in Bezug auf Verkehr, Straßenführung und Oberflächencharakteristika zu analysieren. Die simultane Form des Modells erlaubt die von den Präferenzen der Nutzer abhängende Einbeziehung der Anpassungsmöglichkeiten zwischen Komfort, Geschwindigkeit und Ungewißheit. Die Darstellung ermöglicht insbesondere die Unterscheidung zwischen zwei Komponenten des objektiven (beobachteten) Risikos, nämlich dem mit der Geschwindigkeit verbundenen berechenbaren Risiko und der Ungewißheit im Sinne von Frank Knight. Gleichzeitig wird erstmalig ein neues empirisches Maß des aus der Erwartungsnutzentheorie stammenden Konzepts des "wahrgenommenen Risikos" (*Expected maximum insecurity EMI*) genutzt, das auf die Ungewißheit fokussiert; dies erleichtert die Identifizierung der Bedeutung des wahrgenommenen Risikos unter den Bestimmungsgründen der Geschwindigkeitswahl. Die Schätzstruktur besteht aus drei Gruppen von Gleichungen: Die ersten beiden erklären Unfallhäufigkeiten und ihre Schwere mit diskreten Logit-Modellen, die in einfacher Form die Einbeziehung von Nichtlinearität und endogener Schichtung der Stichprobe ermöglichen. Die Geschwindigkeitsfunktion, die auch nichtlineare Gleichungen enthält, erklärt gleichzeitig den Mittelwert und die Varianz der Geschwindigkeit. Der Ansatz der endogenen Stichprobenerhebung erlaubt schließlich die Vervollständigung der Datenbasis zu niedrigen Kosten, um mögliche Fehler infolge fehlender Beobachtungen, die nicht orthogonal zu den verfügbaren sind, zu korrigieren.

Schlüsselbegriffe: Verkehrsunfälle, Häufigkeit, Schwere, Durchschnittsgeschwindigkeit, Standardfehler der Geschwindigkeit, Mehrfach geschichtete DRAG Struktur, Simultanetät, Frank Knight, Ungewißheit, Risiko, *Expected maximum insecurity EMI*, Infrastruktur, Auslegung, Geometrie, Oberfläche, Straßenführung, Nichtlinearität, Maximum Likelihoodmethode, Box-Cox, Logit, endogene Stichprobenschichtung, TRIO Regressionsprogramm, Wahrscheinlichkeitspunkte, Elastizitäten, Frankreich, MELTT-, SETRA-, SIREDO-Stationen.

Table of contents

Résumé, Abstract, Zusammenfassung i

Table of contents iii

List of the figures and charts v

1. Introduction : characteristics of the pavement and road users' satisfaction 1

2. The particular context of the management system GiRR-SRATÈGE 2
   - 2.1 Measure and expectation of the pavement conditions 2
   - 2.2 Pavement condition, safety and journey time 3
   - 2.3 Interactions between speed and safety 4

3. The behaviour model 4
   - 3.1. Intuitive formulation: distinction between risk and uncertainty 4
   - 3.2. Estimated model 7
     - *A. Perceived risk measurement* 7
     - *B. Speed measurement* 8
     - *C. Data limitations and selected model* 9
     - *D. Data base used* 9

4. Econometric formulation of the system of equations 13
   - 4.1 Non linearity of variables 13
   - 4.2. The road safety equations 13
     - A. Relevance of discrete choice models 13
     - *B. Specification of an unordered and non linear Logit model* 15
     - *C. Endogenous stratification of the sample* 16
   - 4.3. Speed equation 17

5. Results 19
   - 5.1. Available data and tests 19
   - 5.2. Chosen results : safety equations 22
     - *A. Statistical comments* 22
     - *B. Behaviour: the expected speeds* 23
     - *C. Traffic: the percentage of heavy goods vehicles (HGV)* 25
     - *D. Road-dimensions in height and longitudinals: turns and crown* 26
     - *E. Road-dimensions in width: three lanes* 28
     - *F. Surface Road-index: smoothness* 29
     - *G. Surface or structure road-indexes* 30
   - 5.3. Complete results: speed equation 30
     - *A. Statistical comments* 30
     - *B. The mean speed equation* 32
     - *C. The equation of standard deviation of speeds* 33

D. Conclusion 35

6. General conclusion 35

7. References 37

Appendix 1. Description of the 78 available variables 41

Appendix 2. Accident probability and conditional accident severity results 44

Appendix 3. Associated binaries with variables containing zero observations 48

List of the figures and charts

Figure 1. Simulation process of maintenance policies of pavements 2
Figure 2. The simultaneous multilevel frame selected 5
Figure 3. Data base used 10
Figure 4. Individual speeds observed 11
Figure 5. Effect of a variation in mean speed on accident probability 24

Chart 1. Recent discrete choice models explanatory of severity 15
Chart 2. Explanatory speed models 18
Chart 3. Used or selected variables classified by groups 21
Chart 4. Estimation of the frequency and severity: adjustment to the observations 22
Chart 5. Behaviour effects: mean speed and standard deviation of the speed 23
Chart 6. Traffic effects: the percentage of heavy goods vehicles 25
Chart 7. Road effects: three-turn sequence and crown of variable radius 27
Chart 8. Road effects: the horizontal geometric elements 27
Chart 9. Road effects: three-lane road 28
Chart 10. Road effects: smoothness 29
Chart 11. Result of speed equation estimation 31

1. <u>Introduction : characteristics of the pavement and road users' satisfaction</u>

It is a common fact that the characteristics of the design and maintenance of pavements imply different levels of services for road users. The purpose of this study is to combine the procedures of maintenance decisions with the idea of services for users. Most of the pavement management systems already existing are created to maintain structural patrimony of the pavements. However, we notice that the interests in saving time in transport and road safety are more and more important. The present societies require better levels of services from the infrastructures (Conti *et al.*, 1990; OCDE, 1987, OCDE, 1996). This new concern is starting to be taken into consideration by the road managers and the research activities are progressively oriented toward the idea of user's satisfaction.

In the actual pavement management systems, the concern about the levels of services is quite new. Most systems discard the concerns of economical comparison of different maintenance strategies (ex. Kullkarni, 1984; Butt *et al.*, 1994). However, in the case where they clearly take these concerns into account, they only pay a brief attention to them. A common presumption is that a regular structural pavement maintenance is resurfacing the road therefore restoring the surfaces characteristics. Nevertheless, when thousands of vehicles travel on the roads each day, savings, even small ones, can be used to improve the condition of the pavements. The model *Highway Design Maintenance* (Watanada *et al.*, 1987), created by the World Bank for developing countries, is a good example in that field. It clearly incorporates costs funded by users in the economical evaluation of different maintenance strategies. Unfortunately, the established relations can not always apply to the contexts of industrialised countries and prime subjects such as road safety are not integrated in the model. Certain countries, France for instance, tried to combine these concerns with the procedures of decision, yet the evaluation is still imperfect (SETRA-CETE Méditerrannée, 1997).

This study is about the development of pavement management system created by the French administration concerning the level of user's service. It allows evaluating the relations between the characteristics of pavement, the travelling speed and the road safety. The results of the study will be used as a reflection on the construction of an expert rating allowing us to integrate the costs linked to these two subjects in the procedures of choice of maintenance strategies.

Our intention is to study the impact of the characteristics of pavement on the journey time and the road user's safety. To do so, we are launching a behaviour model with simultaneous structure. This structure shows the user's compromises between speed and safety. In addition, the estimation structure suggested completes the low cost data base, which offers interesting perspectives, in a technical angle as well as in the progress of the research in this field.

The next section shows the French pavement maintenance system. It also explains the choice of the two main subjects, journey time and safety, among all the possibilities. The third section displays the structure of the model used in the data base. The fourth section specifies the econometric formulation and the last section comments the main results obtained.

2. The particular context of the management system GiRR-SRATÈGE

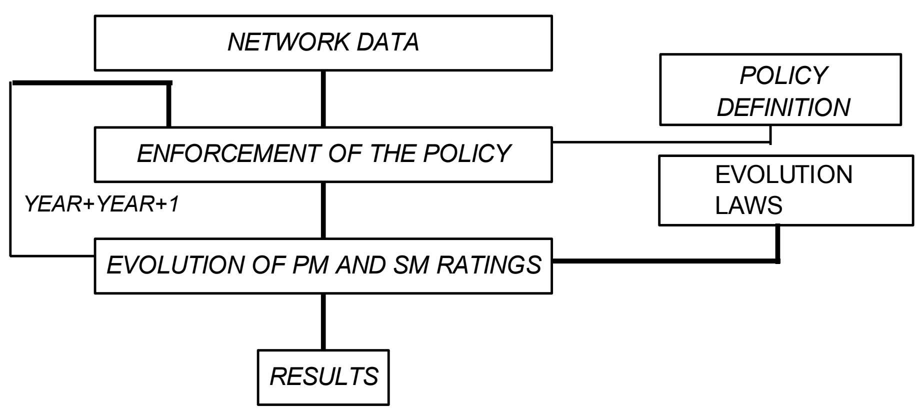
2.1. Measure and expectation of the pavement conditions

Since 1992, in France, the equipment services have developed a pavement control system, the line *GiRR* (*Gestion Intelligente des Réseaux Routiers*), to manage the maintenance of the road network in a more rational way. Data are used to give to each portion or "section of road" a rating of the structural qualities, or a patrimony mark (PM see Figure 1), and a rating of the quality of roadway or surface mark (SF see Figure 1), who together characterise the condition of the pavement.

The patrimony mark is an expert rating set up by elementary accredited ratings related to the damaged structure of the pavement. The value of the mark is directly related to the conventional restoration. This gives an estimation of the cost related to different operations of maintenance. On the other hand, the surface mark refers to the user's service: it is established the same way as the patrimony mark, on a basis of elementary indicators of superficial flaws of roadway.

The *GiRR-STRATÈGE* program enables the simulation of the evolution of a road condition, in accordance with a maintenance policy and a budgetary restriction, and estimate the long term effects. The software incorporates a continual process. Each simulates the evolution of the network during a year. Starting with the present condition, the process applies an evolution law from which we can determine the reliability (Vernier, 1998), established in accordance with the maintenance accomplished. The annual evolution of the network, under different maintenance practices, is obtained by an original product using the chains of Markov process.

Figure 1. Simulation process of maintenance policies of pavements



Several strategies can be estimated, some assuring an improved surface, others concerning essentially the structural characteristics or even combining two pavement conditions, as indicated in figure 1.

The first results of the estimation (SETRA-CETE Méditerranée, 1997) question the relevance of the surface mark. Indeed, if this variable combines several elementary indicators of quality of surface, nothing indicates its relation to the comfort felt by the user, the safety, the maintenance costs of the cars, the journey time, the fuel consumption or the sound of the wheels against the road. So, the evaluation results, established on the basis of an amount of work of restoration, do not take into account the advantages of maintenance for the user. It is therefore important to implicitly include the notion of users' satisfaction in the decision process of the software.

The following study is a first approach to a better definition of this mark, in term of the value of using the infrastructure and not on pavement conditions, to different thematic related to the users' satisfaction. The result of the study constitutes a basis to determine a new surface mark that incorporates the total costs for both the road users and the work master.

2.2. Pavement condition, safety and journey time

Pavement maintenance is aimed to satisfy the users by assuring safety in circulation, regulating distances, permitting an acceptable level of comfort while driving, and by lowering nuisances caused to residents (noise, insecurity, pollution). The field of road managing is particularly concerned by these different thematics. Indeed, the condition of pavements affects these characteristics. A levelled pavement assures a comfortable drive at high speed. In contrast, pavement deterioration (cracks, ruts, repairs) reduces comfort, adds time used for transportation and affects road safety. Many elementary factors play a role in road users' satisfaction and some may even act in opposite direction. For example, smoothness of the pavement has a positive effect on comfort and speed, but if the infrastructure is not designed for that level of quality (conception norms of the geometrical layouts, width of the lanes...), it may on the contrary have a disastrous effect on road safety.

Since an intense study of the entire characteristics (noise, comfort, fuel consumption, safety and journey time) requires lots of data base, we have selected these two particular subjects: safety and journey time. For many reasons, these two elements seemed to stand out the most. Increased traffic and travelling restraints require the best possible service in terms of journey time. The latter has been the subject of many studies (Watanatada *et al.*, 1987; Planco Consulting, 1993). The costs related to degradation of this level of service seemed very high, considering the methods of time evaluation. Planco Consulting determines the cost of journey time as a cause of materials cost proportional to time. He accomplishes a proportion between time and capital invested in vehicles, purchase, depreciation and maintenance.

Road safety is also a major preoccupation. Car crashes are indeed a main cause of the death rate in industrialised countries; in 1998, we find a bit less than 100 000 deaths in the 15 countries of European Union and North America. World-wide, there are approximately 17 millions people killed by automobiles since the beginning of the century (Bergeron, 1999).

2.3 Interactions between speed and safety

Another point in favour of the choice of the two main factors is their interaction. We generally associate the level of road safety with the driven speed, and vice-versa. Moreover, we often consider that journey time and safety work in opposite direction. There is a general consensus that when both are on an equal level, accident severity increases with speed and the accident probability increases with deviations from the average traffic speed. It is important to determine explicitly the relations between those two elements and to find the best compromise between the chosen speed and user's safety.

Despite the general agreement that this interaction exists, we could not find any literature on the subject: we found models where the safety level depends on speed and where speed is explained by another equation, precisely recurring models, but no model where risk retroaction awaited or "perceived" on the selected speed is simultaneously included in the model. For an example, two levelled models (Cardoso, 1997 et Aljani *et al*. 1998), or three levelled models (Jaeger, 1998), are recurrent.

Inspired by random utility theory we formulate a measure of the felt risk which will depend on the risk associated with speed on one hand and with pavements characteristics (and in our model, with traffic) on the other hand. To explain them, we will associate these components to the "risk", which means calculable risk, and to the uncertainty, or if you prefer non-calculable risk, as described in Knight's book (1921).

As far as the part played by pavements characteristics is concerned, it can be found in many ways in the explanatory equation of the safety level or the speed found in the literature, and without characterising it, with the many available elements. Barber *et al*. (1998), for example, using the Zeeger *et al*. (1991) model, explain the frequency of accidents in curbs with the five elements model such as traffic and four geometrical characteristics of curbs.

3. The behaviour model

3.1. Intuitive formulation: distinction between risk and uncertainty

The goal of producing a model being to evaluate the user's satisfaction in relation to speed and safety, the specification must allow the evaluation of the users' preferences and the impact of the road characteristics on the associated level of satisfaction.

Multilevel frame. Taking into account the perception of the road environment, the user will choose a speed level and feel a risk level, without necessarily controlling the consequences. These two evaluations depend on each other, and the structure of the general model must take it into consideration, as indicated in figure 2, where we find a structure D - P (Demand-Performance) of multilevel type DRAG (Gaudry, 1984) simplified and specified. On the one hand, the characteristics of M-C vehicles being given, the entire activities of self-protection and selfassurance are reduced to speed choice - we do not show the condition of the vehicle and the use of seatbelt. On the other hand, the speed chosen **V** depends on the perceived risk defined as expected maximum insecurity EMI ≡ u(**A,G**): the perceived risk does not directly depend on the entire N-I characteristics of the road environment. Only the characteristics k(N-I), associated with the driving comfort chosen, play a direct role in the selected speed. Besides, we consider as exogenous this level (unobserved) of comfort determined with the selected speed which determines a specified risk called "calculated risk".

Risk and uncertainty. In this frame, where the simultaneity between the objective risk (observed) and the speed is clear, we add an important precision by defining the hidden variable u(**DR**, N-I), M-C as a generator of *residual* uncertainties conditional to the traffic, the road and the vehicle, after adjustment of the speed **V** by a driver of general characteristics Y-G having a preference for risk and giving a value to time. This is the application of Knight's (1921) idea to a driver, where we distinguish between the part of the risk that can be calculated and the one that can not.

This uncertainty indicates the surprise element shown by notions such as "road legibility", where the driver undergoes an objective risk different than the one that is measurable or predictable by his choice of speed. Without this "surprise", the objective frequency or severity risk, **A** or **G,** would represent for this individual the risk explained by speed. This uncertainty will make way for intervention of the road safety authorities, for example by making useful the information on the condition of the road environment. The intervention should lessen the surprise element and allow the driver to control his risk. If there was never any surprise, the speed selected would then explain the observed risk. The latter would imply a level of perceived risk sustaining exactly that chosen speed.

Figure 2. The simultaneous multilevel frame selected

(Non estimated equation)	Road demand **(DR)**	←[(––, ––––), ––––,––––– ,––––– ,––- , ETC.]	**Exposure risk**	**D**
VICTIMS ←	Frequency and severity **(A) , (G)**	←{**Calculated risk**, Uncertainty [**traffic** + road + vehicle] Driver, ETC.} ←{(f(**V**)), (u[(**DR**, N-I), M-C]), (h (Y-G)) }	**Frequency and severity risks**	**P**
Travel Time ←	Realised speed **(V)**	←{**Perceived risk**, Driving comfort, Driver, ETC} ←{(EMI ≡ u(**A,G**)), (k(N-I)), (M-C), (Y-G) . }	**Calculated risk**	

Uncertainty does not mean unawareness by means of perceived risk. This corresponds to Knight's idea to whom uncertainty has vaguely Bayesian characteristics. Uncertainty is not exactly explainable, but it is in a global way far more comprehensible by intuition than in a specific way ("direct" in his language): we are talking of "dangerousness" as this *part* of perceived risk which relies on uncertainty and we say that the driver perceives the global risk or dangerousness of the driving environment without measuring the specific risk related to each element of that environment. We will see shortly that our measure of perceived risk will here be centred *de facto* on that uncertainty, excluding other factors for practical reasons of information availability and of econometric formulation.

Theoretical model. We therefore propose a simultaneous structure containing three equations. The first two equations are related to the accomplishment of the road safety for an individual and the last equation, to the speed performed by the road user. The specifications chosen for the first two are multinomial discrete choice Logit models but we only specify later on in 4.2.B and in 4.3. the mathematical formulation of the three equations forming the system.

$$\text{Frequency :} \quad y_a^* = \beta_{1,a} R_a + \beta_{2,a} YG_a + \beta_{3,a} MC_a + \beta_{4,a} y_v + \eta_a \tag{1}$$

$$\text{Severity :} \quad y_g^* = \beta_{1,g} R_g + \beta_{2,g} YG_g + \beta_{3,g} MC_g + \beta_{4,g} y_v + \eta_g \tag{2}$$

$$\text{Speed :} \quad y_v = \beta_{1,v} R_v + \beta_{2,v} YG_v + \beta_{3,v} MC_v + \beta_{4,v} y_a^P + \beta_{5,v} y_g^P + \eta_v \tag{3}$$

With

y_a^* and y_g^*	:	vectors of unobservable latent variables representing the risk for a given individual. To each vector corresponds a observed discrete variable taking different values according to the levels of y_a^* et y_g^*;
η_a and η_g	:	perturbations according to a given distribution ;
R, Y-G, M-C	:	vectors of explanatory variables characterising the road taken (R= DR, N-I), the characteristics of the driver (Y-G) and of the vehicle (M-C);
y_v	:	continuous observed variable describing the speed adopted by the road user;
η_v	:	perturbation associated to the speed equation, of zero mean and a standard deviation $\sigma^2 I$;
y_a^P and y_g^P	:	variables describing the risks perceived by the road user in terms of frequency and severity;
$\beta_{i,a}$, $\beta_{i,g}$ and $\beta_{i,v}$	:	vectors of parameters to be estimated.

This system considers the dependencies that exist between the speed performed and the risk perceived on the road (y_a^P et y_g^P). So, at a high level of perceived risk, we will find a lower speed than at a lower level of perceived risk. The main difficulty is to characterise the perceived risk that is unobservable. The Delhomme and Malaterre's (1990) bibliographic analysis shows that the perceived risk is often evaluated on an arbitrary scale, a measurement excluded in our case. However, there might be other possibilities.

3.2. Estimated model

A. Perceived risk measurement

Taken from discrete data ? The first possibility is to consider that the perceived risk is characterised by the realisation of the associated discrete variable. So, the speed behaviour is described as a function of the road characteristics and accident occurrences in terms of probability and severity. However, since accidents are rare, it is unlikely that individuals will pay any particular attention to them or see them as a perceived risk.

As linear function of road variables ? It seems more realistic to assume that an individual associates a level of risk with each site, whatever the history of that site is in terms of road safety, considering for example the frequency or severity perceived risks as linear functions f_a and f_g of the road characteristics R. This assumption is reasonable because it lies on the analysis of the link section configuration and the traffic, and is an acceptable evaluation of the perceived risk. This measurement would then transform (3) in :

$$y_v = \beta^*_1 R_v + \beta^*_2 YG_v + \beta^*_3 MC_v + \eta_v, \tag{3*}$$

where, because of the linearity fact, the coefficients combine the already distinct effects, so that the specific part played by the perceived risk is no more identifiable. No correction is then done in the speed equation if the measure of perceived risk is a linear function of explanatory variables : this possibility is also to be excluded.

Expected maximum utility. We would rather define the perceived road risk as expected maximum insecurity EMI, identical to expected maximum utility EMU constructed from the inclusive value of Williams-McFadden (Williams, 1977) :

$$y^p_a = I_a(R, YG, MC, y_v) = \left(\sum_{i=1}^{M_a} exp\left(V^a_i\right) \right) \tag{4-A}$$

$$y^p_g = I_g(R, YG, MC, y_v) = \left(\sum_{i=1}^{M_g} exp\left(V^g_i\right) \right) \tag{5-A}$$

where V_i^a (*resp.* V_i^g) represents the utility associated with the i^{th} realisation in the model of accident probability (*resp.* for accident severity) and is related to the characteristics R, Y-G and M-C of the section, the driver, the vehicle and the chosen speed y_v; moreover, M_a (*resp.* M_g) means the number of possible realisations of accident probability (*resp.* for severity). The natural logarithm of the inclusive values $I_a(R, YG, MC, y_v)$ and $I_g(R, YG, MC, y_v)$ in this case represents the expected maximum insecurity (EMI) associated with the two models of road risk, the expected maximum that the user can expect considering what he can do.

Such measurements, called expected maximum utility (EMU), are often used in the applications for transportation demand, where the hierarchic problems are frequent (Ben-Akiva and Lerman, 1985; Ortuzar and Willumsen, 1990). These measurements have the benefit of characterising the perceived risk by a unique variable in each modality of the road risk (accidents probability and severity), where the accident probability is explained by a two-alternative model and the severity by a three- alternative model. A characterisation of the perceived risk by the different realisations of risk would have needed taking into account, in the speed model, 6 additional explanatory variables representing the different possible combinations of the frequency and severity.

A first-step estimator. As first-step estimator of the perceived risk defined in (4-A) and (5-A), we adopt the two following multinomial Logit:

$$y_a^* = R_a\beta_{1,a} + YG_a\beta_{2,a} + MC_a\beta_{3,a} + \varepsilon_a \tag{4-B}$$

$$y_g^* = R_g\beta_{1,g} + YG_g\beta_{2,g} + MC_g\beta_{3,g} + \varepsilon_g \tag{5-B}$$

With $E\left[\varepsilon/R,\ YG,\ MC\right] = 0$ and $E\left[\varepsilon_a\varepsilon_g/R,\ YG,\ MC\right] = 0$.

These two measurements, defined by exogenous variables, are asymptotically independent of errors. The values $\hat{y}_a^*$ and $\hat{y}_g^*$ calculated by (4-B) and (5-B) are therefore a first-step estimate, or instrumental, of a measured perceived risk where all the variables of (1) and (2), the chosen speed included, are accounted for in (4-A) and (5-A). A second-step would then include in (4-B) and (5-B) not only the endogenous already present but also a first-step estimate of the chosen speed, $\hat{y}_v$. The limitations of our speed sample, however, urge us to consider some estimates of second-step only in the event of future studies.

B. Speed measurement

Simultaneity bias and independent samples. If these choices of (4-B) and (5-B) to describe the perceived risk remove a "degree" of simultaneity, the speed remains an explained endogenous variable as well as explanatory, causing a possible error (bias) in the estimate of (1) and (2). The problem still persists even if, data not being computed on the individual speed of the accident victims, we have to use the derived measurement of the mean speeds observed on these particular sections - formulation of its own interest - as the mean speed and standard deviation of the individual speeds. Another difficulty is that these measurements only exist for 17 sections. So, we generate expected speed measurement from equations (8) and (11) estimated for these 17 sections and we will use them in equations (1)-(2) excluding the 17 sections of the sample used for the estimation of these two equations. Indeed, if the sample used to estimate the speed equation is different than the one used to estimate the frequency and severity equations, it is known (Arellano et Meghir, 1992) that the measurement $y_v^e = E\left(y_v\right)$ therefore deriving and replacing the speed observed in the equations (1)-(2) will differ from the errors of theses two equations, and will produce consistent estimations.

C. Data limitations and selected model

Use of disjoint samples is made easier because the data on 60 000 observations of instantaneous speed that we have at our disposal, taken on 17 road sections, is considered as random. The low number of sections limits the quality of the measurement found for the other sections.

Estimated Model. Moreover, no data base related to individual or vehicle having been computed, we have to estimate a model (1)-(5) cut down, with explicatory endogenous derived in two preliminary steps.

Probability : $$y_a^* = \beta_{1,a} R_a + \beta_{2,a} \tilde{y}_v^e + \eta_a \quad (6)$$

Severity : $$y_g^* = \beta_{1,g} R_g + \beta_{2,g} \tilde{y}_v^e + \eta_g \quad (7)$$

Speed : $$y_v = \beta_{1,v} R_v + \beta_{2,v} \hat{y}_a^P + \beta_{3,v} \hat{y}_g^P + \eta_v \quad (8)$$

where $\hat{y}_a^P$ and $\hat{y}_g^P$, drawn from first-step models describing the perceived risk associated with accident probability:

$$y_a^p = I_a(R) = \left(\sum_{i=1}^{M_a} \exp\left(V_i^a\right) \right), \quad (9)$$

as well as perceived risk associated with accident severity:

$$y_g^p = I_g(R) = \left(\sum_{i=1}^{M_g} \exp\left(V_i^g\right) \right), \quad (10)$$

and where $\tilde{y}_v^e$, the expected speed, is then calculated from (8) :

$$y_v^e = E\left(y_v\right) \ ; \quad (11)$$

with finally :

$$E\left[I_a(R)/\eta_v\right] = E\left[I_g(R)/\eta_v\right] = 0 \text{, as well as } E\left[y_v^e/\eta_a\right] = E\left[y_v^e/\eta_g\right] = 0 . \quad (12)$$

In this new disjoint sample system, y_v is independent of (y_a, y_g): in consequence, the equation (8) generates expected speed measurements, $\tilde{y}_v^e$, for each section considered in the equations of accident probability and severity, that will allow a consistent estimate of the equations (6)-(7) because an eventual correlation between errors η_a and η_g do not affect the consistent estimates of second-step.

D. Data base used

This model has been applied to a three different sources data base, as indicated in Figure 3, where we find that the characteristics of drivers Y-G and of vehicles M-C are available only for the damaged vehicles, not for the individuals without accidents or for the vehicles for which speed has been monitored. The three source-bases are grouped thanks to a common marker system: for example, the accident occurred at point A is at 700 m. from marker # 3 or at 3300 m. from marker # 0.

Figure 3. Data base used

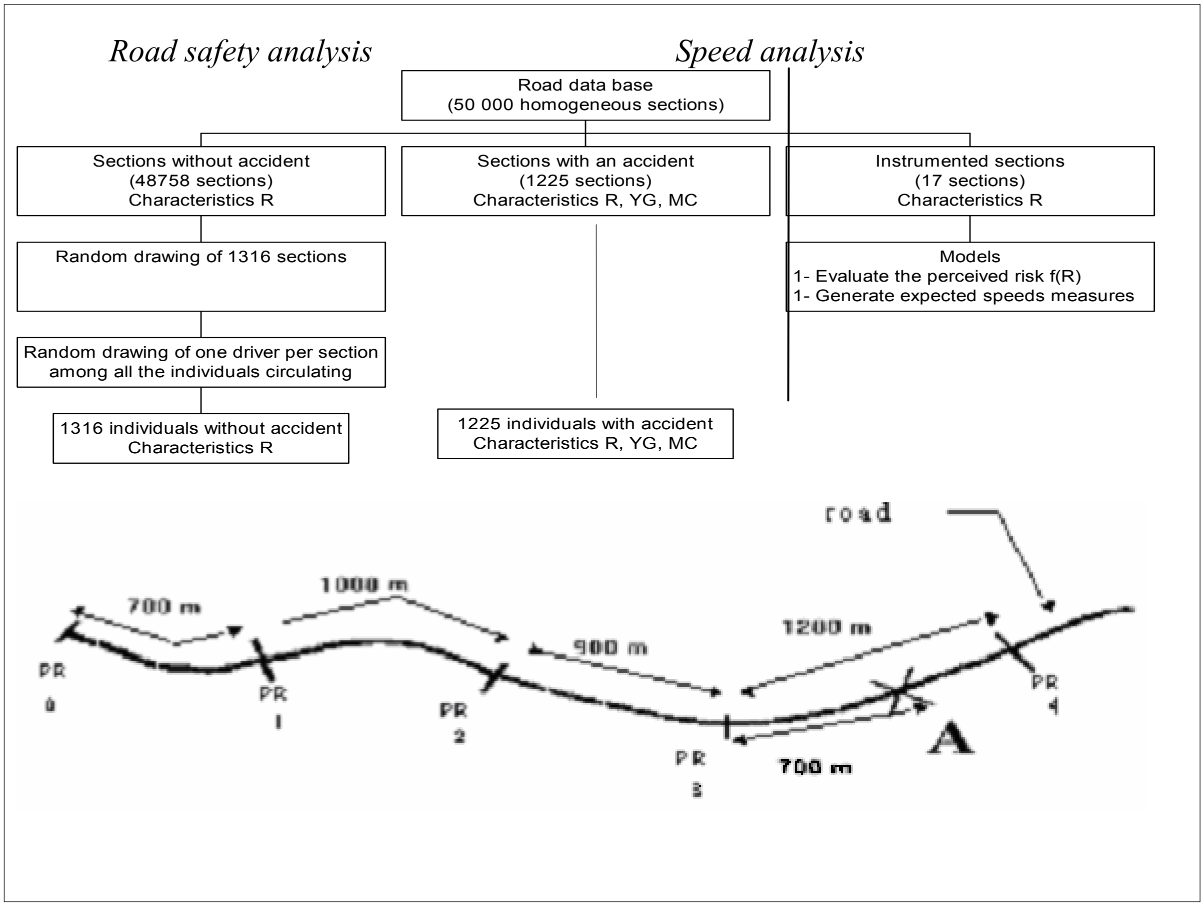


Figure 4. Individual speeds observed

Figure 4.1. Individual speeds in terms of length of the vehicles, the 17 stations (complete base)

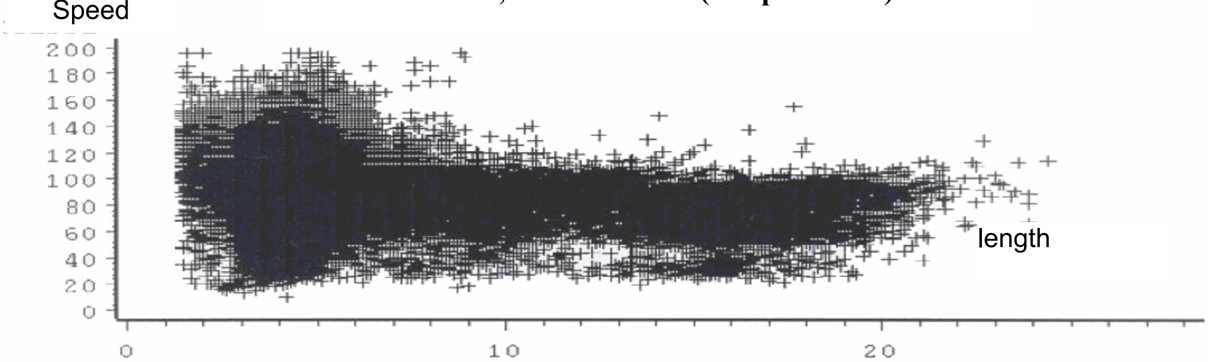


Figure A4.2. La Brillante Station, dep. 04

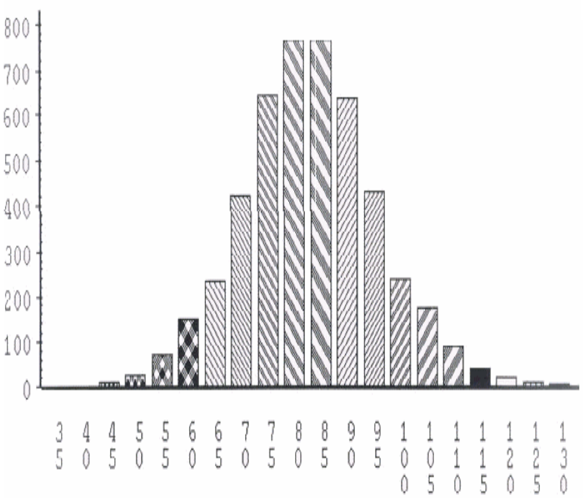


Figure A4.3. Station of Tenay, dep. 01

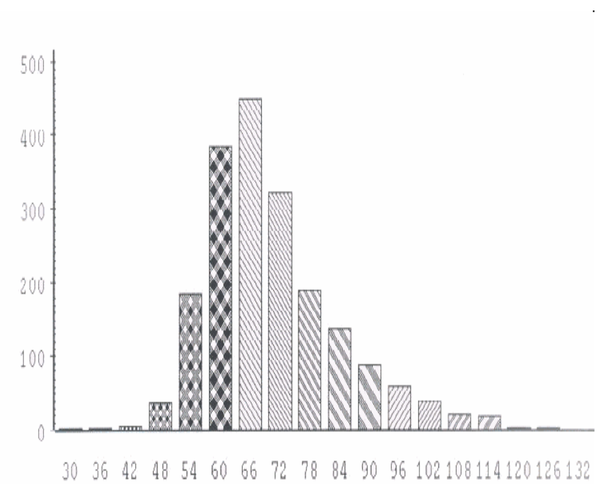


Figure 4.4. Station of Coetmieux, dep. 22

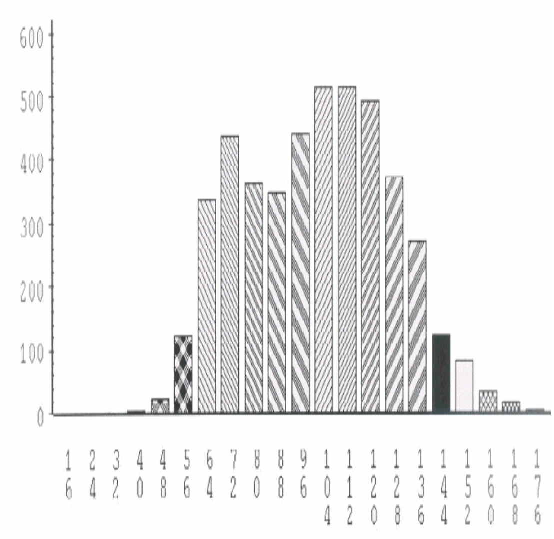
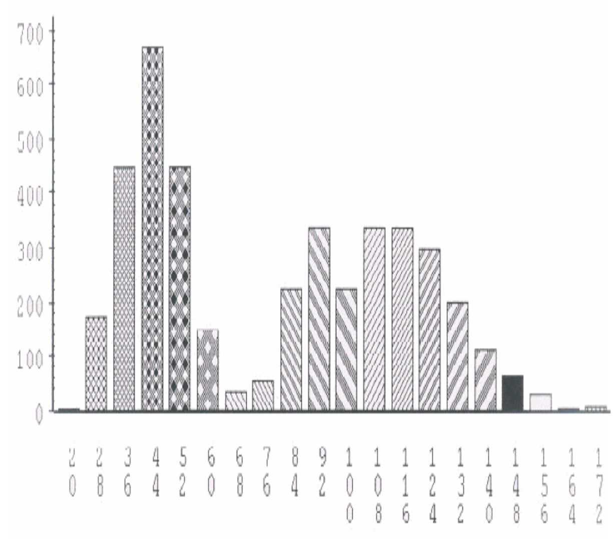


Figure 4.5. Station of Plelan le Petit, dép. 22



Accident data. The accident data base covers a five-year period (1991-1995). Among all the available information, we have only worked with the one related to the seriousness of the injuries per accident and to the road location.

Road data. The road data base is the main one. It is composed of 50 000 sections of national roads of varying length and describes the characteristics of pavement surface and structure as well as geometry and traffic characteristics.

Concerning the *traffic*, we have used the file *" histo "* available at SETRA where are registered the annual daily traffic averages (ADA) for light vehicles and heavy goods vehicles. These values come from the measurement stations found on the national French network and are similar for average length sections of 25 kilometres.

The overall measurements related to the pavement condition are from the year 1993 and comes from the survey campaign of *Image Qualité du Réseau National (IQRN - Quality Picture of the National Roads)*. The measurements are available for the entire French national network and characterise the sections at the walking pace of 200 meters. These measurements, from where are taken the variables of groups 14 and 15 of the Annexe 1, are elementary indicators of road degradation describing the condition of both the pavement *structure and surface*. The structural conditions determine the pavement resistance to negative climate and traffic effects. The surface conditions are responsible for the riding comfort, safety and, in general, the level of service offered to road users.

As for the *geometry* data, the measurements are obtained through the gyroscopic stations. These measurements underwent validation from the Public Work department for the entire sections of the research. The network was therefore divided in homogeneous geometric zones according to the sections horizontal surface or in length (turn, straight alignment, downward and upward sloping road…) and its difficulty (turning radius, slope percentage). Moreover, we have created more variables characterising the approach zone of the sections according to the horizontal layout and the longitudinal profile. These approach zones describe the sequence of geometrical elements on a 450 meters length preceding the current section.

Speed data. Finally, the *speed* data base illustrated in Figure 4 contains 60 000 spot speed measurements especially obtained for this study on a unexceptional 24-hour time period and on 17 of the sections in the road data base. It was impossible to measure the speeds on the overall sections studied: the counting stations *(SIREDO),* capable of registering vehicle spot speed, are not distributed throughout the entire territory and the radar recording methods seemed too expensive since we use 2 541 sections in the accident frequency model. The measurements have been realised by fixed counting stations, installed by the Public Works services and used for annual traffic level computing : their location is therefore not relevant to our research. The selected sections on which speed has been measured are considered as random.

4. Econometric formulation of the system of equations

4.1 Non linearity of variables

One of the authors (Gaudry, 1998) has recently summarised the theoretical and practical importance of the non linearity, in the Logit models as well as in the typical regression models. The non linearity is generally more reasonable than the linearity. It gives a better adjustment to the data and influences the elasticity and even the *signs* of the regression. These considerations and the experience explain that the Box and Cox (1964), the most commonly used non linear transformations in econometrics, because of the parameter λ applied to the continued and strictly positive variable X and defined by

$$X^{(\lambda)} = \frac{X^{\lambda} - 1}{\lambda} \quad , \text{if } \lambda \neq 0 \qquad (13)$$

$$X^{(\lambda)} = ln(X) \quad , \text{if } \lambda = 0 \qquad (14)$$

contains as particular case the linear specifications (λ=1) and logarithmic (λ=0).

In the case of a variable that would contain zero observations, the TRIO software package that we use (Gaudry *et al*., 1995) authorises the transformation and creates an auxiliary and compensatory variable that preserves the invariance of the forms in the X measurement units, as explained in Annexe 3.

4.2. The road safety equations

The severity of accident probability equation is concerned with its frequency, no matter how severe the injury is. In the second equation, the severity is explained according to three levels: (i) the fatal accidents in which at least one person dies ; (ii) the accidents with severely injured people in which at least one person has severe injuries but no one is killed ; (iii) the accidents with slightly injured people in which at least one person suffers from slight injuries but no one is severely injured or killed.

A. Relevance of discrete choice models

The conventional models of road safety that use the individual data are the measurement models (Poisson or negative binomial). For many reasons, we have chosen to opt rather for discrete models very often used in the transport mode choice modelling.

First of all, we chose these models because of the type of data used in our modelling. We evaluate each individual separately and we observe his accident probability, and severity in case of an accident. The frequency model can only then rely on the binary variables (whether or not the individual has an accident). Now, we know that the counting models are particularly recommended when the tail areas of a distribution are well represented. The main application of these models is for the explanation of the number of accidents, when the observed values can not justify a Gaussian-type error.

Besides, the severity model relies on a type of accident and, as their name indicates, counting variables are not category variables. Therefore it would be inappropriate to use a counting model for the modelling of the accident probability or severity, as we have specified here, in spite of the possibility of treating the random heterogeneous and spatial correlation problems (Bolduc and Bonin, 1999) offered by some multinomial distributions.

Frequency. Boyer *et al.* (1992) have shown, in the case of a model where the implication of the individuals has been clearly proven, that the results of the multinomial Logit model are different from the counting models. This could be due to the fact that the observations are not independent in time: the two discrete choice models do not depend on the time independence theory and do not assume that frequency variable is equal to its mathematical expectation, contrary to the Poisson model. The discrete models seem to be more credible, especially if we consider, as in the Logit, the non linearity and the presence of stratified samples. McCarthy and Madanat (1994) are expecting that these new models will improve the quality of the safety models.

Severity. These discrete choice models are classified in two categories: the ordered and the unordered models. This choice is possible for the non binary severity models. An ordered model implies that the severity is classified in order of importance. There are, of course, many interpretations on the severity. The classification in this case can be very touchy and implies a detailed knowledge of the injuries. Ranking the degrees of severity can be a risky task and we cannot certify, *a priori*, that the new categories will follow an order of severity. Furthermore, Amemiya (1985) stipulates that an ordered model must be used with caution. Indeed, on the assumption that the real model is unordered, an ordered model can lead to probability estimates bias; when the opposite is wrong, we then only find a loss of efficiency.

We have therefore chosen an accident severity model based on the non linear multinomial Logit, just as others did with linear models, as shown in Chart 1. Weiss (1992), in a study on accident severity, has pointed out that these studies have limitations due to the lack of behavioural variables, such as speed. Without these variables the models used are absolutely converted into mechanical formulas, and the structural equations into reduced forms: consequently, (3*) is a reduced form of the speed equation.

Chart 1. Recent discrete choice models explanatory of severity

Linear models	ATO EGUAKUN and WILSON 1995	O'DONNEL and CONNOR 1996	WEISS 1992	SHANKAR *et al.* 1995, 1996
Model type	Ordered Probit		In-box Logit	
Dependent variable	Severity level			
Number of levels treated	(?)	4	6	4
Classes and explanatory variables				
Behaviour				
speed		1	1	1
seatbelt/helmet	1	1	1	1
Traffic				
type collision		1	2	4
Road-Pavement				
geometry				2
Road-Environment				
meteorology				2
collision hour		2	1	1
Vehicle				
vehicle type		3		2
position (injured)		1		
Driver				
age	1	2	1	1
sex	1	1		1
experience			1	
alcohol		1	1	1
Number of observations	2 500	18 069	770	1 505

B. Specification of an unordered and non linear Logit model

Direct choice of speed and objective risk. The description of the behaviours found in 3.1 suggests that preferences are represented by a utility function, where time utility and riding comfort depend on the risk chosen by the driver, in relationship with the infrastructure on which he his travelling on and his chosen speed (Underwood *et al*., 1993). Utility will depend on the evaluation of an anticipated loss associated with the perceived risk. Its maximisation will determine a speed choice, an object of risk control chosen by the individual and an important factor of the objective risk (described as accident probability and severity) to which he is exposed.

The realisation of the objective risk independence. The largely independent characteristics of the objective risk makes it possible to establish that the residuals ε_i of equations (6) and (7) are independent and identically distributed according to the law of the "extreme values" type of distribution function:

$$F(\varepsilon_i) = P(\varepsilon_i \leq \varepsilon) = \exp(\exp(-\varepsilon_i)). \tag{15}$$

We can demonstrate that

$$\mathrm{P}(\mathrm{Y_i} = \delta) = \frac{exp(\mathrm{V_{i\delta}})}{\sum_{j=1}^{N} exp(\mathrm{V_{j\delta}})} \quad \text{and} \quad P(Z_i = \alpha / Y_i = \delta) = \frac{\exp(V_{i\alpha})}{\sum_{j=1}^{M} \exp(V_{j\alpha})}, \tag{16}$$

where $\mathrm{Y_i}$ is the realisation of the accident probability loss, $\mathrm{Z_i}$ the realisation of severity loss, $\mathrm{V_{i\delta}}$ the utility function associated with the realisation δ for the person i and, if λ_k represents the Box-Cox transformation associated with the parameter $\mathrm{X_k}$, it is expressed as :

$$\mathrm{V_{i\delta}} = \sum_{k=1}^{K} \beta_k \mathrm{X}_{i\delta k}^{(\lambda_k)}. \tag{17}$$

We made sure that the assumption of independence of alternatives, or the calculation with axiom IIA (" Independence of Irrelevant Alternatives ") used by (16), was valid by applying the Small and Hsiao (*cf.* Ben-Akiva and Lerman, 1985) test. This assumption, generally rejected in the case of transport mode choice models, is naturally less incongruous here.

C. Endogenous stratification of the sample

Considering the scarceness and the randomness of an accident, we have made a stratification of the sample. So, on sections where no accident has been recorded, we take into account a single individual among the thousands of travellers who drove by. On sections where we have recorded an accident, we only select the driver that has been victim of an accident. This drawing of a sample is often used in studies concerning transport mode choices and also in biometrics. What makes the segmentation of samples so appealing is the fact that setting up data bases is a lot less expensive. In addition, the estimation techniques linked to such samples are perfectly known and operational : Manski and Lerman (1977) conducted an analysis of the stratified sampling processes by means of the maximum likelihood estimator in a more general situation than the random case.

In the discrete choice models, we can formalise the population distribution on both the characteristics of individuals *X* and the alternatives *i*. Note *f(i, X)* the combined distribution of *i* and *X*. So, *f(i, X)* can be factorised as a product of a conditional distribution and marginal distribution :

$$f(i,x) = p(i / x, \beta, \lambda) p(x) \tag{18}$$

In the case of a random sample, we find:

$$L = \prod_{n=1}^{N} \prod_{i=1}^{I} f(i, x_n)^{y_{in}} = \prod_{n=1}^{N} \prod_{i=1}^{I} p(i / x_n, \beta, \lambda)^{y_{in}} p(x_n)^{y_{in}} \quad (19)$$

with $p(x_n)$ independent from β and λ . This probability therefore does not interfere with the calculations of estimators. When the random sample is stratified according to G groups, the following likelihood is obtained, for the entire sample:

$$L = \prod_{g=1}^{G} \prod_{n=1}^{Ng} \prod_{i=1}^{I} \left(\frac{f(i, x_n) H_g(i)}{W_g(i)} \right)^{y_{in}} , \quad (20)$$

with $H_g(i)$ representing the proportion of individuals choosing the alternative *i* in the total sample and $W_g(i)$ representing the comparable proportion in the stratified sample. Manski and Lerman (1977) have proven, without any loss of generality for a model where $\lambda^* = 1$, that for a sample that is endogenous or stratified by choice, the solution $\pi^* = (\beta^*, \lambda^*)'$ of

$$max_{\pi} \left\{ \sum_{g=1}^{G} \sum_{n=1}^{Ng} \sum_{i=1}^{I} y_{in} \left(\frac{W_g(i)}{H_g(i)} \right) ln(p(i / x_n, \pi)) \right\} \quad (21)$$

is a consistent estimator of $\pi = (\beta, \lambda)'$. We have used this procedure, available in TRIO (Liem and Gaudry, 1993), consisting in weighting each observation by $W_g(i)/H_g(i)$ and named *Weighted Endogenous Sample Maximum Likelihood* (*WESML*).

4.3. Speed equation

The bibliographic analysis outlined in Chart 2 shows two main continuous speed measurements: the mean speed VM and the standard deviation of speeds, linked to V85, the 85^{th} percentile averages of speeds realised on the sections. Other authors study particular cases : for example, Kang (1998) analyses, with the help of an ordered Probit, exceeding speed limits. If it is true that only the mean speed is to be considered when evaluating travel time, the standard deviation has a considerable impact on safety. The speed errors can be, to some extent, controlled by statute laws or a better management of some of the road parameters. We notice, for example, that speed restriction measurements do not have a significant impact on the mean speed, but contributes to drops in high speeds.

The equation (8) defining the expected speed on each section can thus be rewritten as followed:

$$VM_{VM} = \beta_{1,VM} R_{VM} + \beta_{2,VM} y_a^P + \beta_{3,VM} y_g^P + \eta_{VM} \quad (8a)$$

Chart 2. Explanatory speed models

	Collins and Krammes (1996)	Gambard (1986)	Badeau *et al.* (1997)	Lamm *et al.* (1987, 90)	Cardoso (1997)	Godlewski (1985)	Jaeger (1998)
Dependent Var.	V85	VM,V85	VM	V85	V15,V85	V85	VM
Classes and groups or explanatory variables							
Behaviour							
Use of seatbelt							X
Velocity of approach					X		
Traffic							
Kilometre-vehicle							X
Traffic	X			X			
Industrial activity							X
Calendar							X
Road-Dimensions in height and lengthwise							
Degree of curvature	X			X	X		
Turning radius	X	X	X		X		
Curve length	X	NS		NS	X		
Deflexion angle	X						
Degree of slope		X	X	NS	X		
Road-Dimensions in width and length of section							
Ramp length		NS			X		
Lateral obstacle		NS					
Visibility distance		X		NS			
Superelevation		NS	NS	NS			
Shoulder width				X	X		
Road type	X	X					X
Pavement width	NS	X		X	X		
Road-Surface indexes							
Pavement coating type	NS					X*	
Road-Environment and Information							
Meteorology		X		NS			
Signage		NS					
Road-Regulation							
Safety regulations							X
Speed limit	X	NS	X	X			X
Vehicle							
Power							X
Fuel and veh. Costs							X
Driver							
Alcohol							X

* To clearly show the impact of the smoothness level on the speed of vehicles, the selected sites are homogenous in geometry and in traffic. **NS** means " non significant ".

$$ET_{ET} = \beta_{1,ET} R_{ET} + \beta_{2,ET} y_a^P + \beta_{3,ET} y_g^P + \eta_{ET} \quad (8b)$$

where *VM* refers to mean speed and *ET* to standard deviation of speeds.

The econometric specification used to estimate each of these equations by the maximum likelihood method is :

$$y_t^{(\lambda_y)} = \sum_{k=1}^{K} \beta_k \cdot X_{kt}^{(\lambda_{X_k})} + u_t \quad . \quad (22)$$

We tried to use the sample consisting of 60 000 spot speed measurements. However, since these sections are related to no more than 17 sections, we only have at our disposal 17 distinct sets of pavement characteristics to explain these 60 000 measurements. Besides, the vehicle-related information recorded by the SIREDO stations cannot be used in this modelling because it is not available for all the individuals considered in the system. Because of these constraints, we chose to consider the solution of aggregating spot measurements. Because we did not weight these observations, it produces an overestimate of the standard deviations, but does not create any bias of the coefficients.

No distinction, according to the type of vehicle or the traffic conditions, is made. Indeed, it would have been possible to calculate the aggregated speed measurements for: (i) vehicles less than 7 metres in length and longer vehicles (this threshold is clearly shown in this descriptive analysis of Figure 4.1) and (ii) the vehicles travelling by night or by day, as well as (iii) the crossing of the two previous stratification. Nevertheless, the data base on the individuals considered in the road safety analysis contains no precise indication of the traffic composition nor of its time distribution. Therefore, taking into account these stratified measurements generates a risk of introducing observation errors due to the lack of information in the sample associated with the road safety analysis. The aggregated measurements, all vehicles and all traffic conditions combined, are more imprecise but less disputable values.

5. Results

5.1. Available data and tests

Variables and chosen forms. Because we can not anticipate the effect of the different variables, we have undergone hundreds of tests of inclusion χ^2 and of the functional forms of the variables taken in group or individually. In Appendix 1, the variables with a clear definition for the 78 sections describing the traffic and the road observed will be found. They are regrouped in Chart 3 that gives a qualitative summary of the results obtained. We find that similar groups of variables do not explain the frequency or the severity, or the standard deviation of speeds.

Confidentiality. The Service d'Études Technique des Routes et Autoroutes (*SETRA – Roads and Motorways Engineering Department*), which provided the entire data bases, wishes to retain the exclusivity on certain results obtained on 9 of the 10 variables selected (amongst the

26 tested) pertaining to the pavement surface characteristics in the road safety models. It is thus impossible for us to present all the results.

However, we can give the results relating to road surface "smoothness" and the other variables, such as geometry, traffic and behaviour. The findings on safety displayed in Annexe 2 are therefore complete in a numeric sense, but it is impossible to recognise the concealed identity of 9 variables and of binary variables sometimes associated with them. Speeds results in Chart 11 are upright, meaning complete and interpretable.

Results expression. These charts, complete or incomplete, are taken from Tablex deliveries of Trio software package that provides the main estimation results in three separate parts :

I. includes elasticity, evaluated to the average sampling of variables, as well as Student's tests conditional to the betas estimated and therefore invariable to the measurement units of X (Schlesselman, 1971).

   In the accident probability models, we use as elasticity the "**probability points**" measurement:

$$\pi\left(P_n(i), X_j\right) = \frac{\partial\, P(y_n = i)}{\partial\, X_j} \cdot X_j \tag{23}$$

   which is understood as the effect on probability $P(y_n=i)$, effect expressed in probability points, of a percentage variation of explanatory X_j. In the speed model, we use the usual motion of **elasticity**:

$$\eta\left(y, X_k\right) = \frac{\partial y}{\partial X_k} \cdot \frac{X_k}{y} \;; \tag{24}$$

II. is related to non linear transformations included in the model. For each Box-Cox transformation, two Student's *t* statistics are proposed: one to test the null linear hypothesis and the other in a logarithmic form. Our *t* tests are presented for an indicative purpose: computed with the Berndt *et al.* (1974) first derivative method, they are not as precise as the exact χ^2 tests that have guided us;

III. finally shows other results pertaining to the estimation, namely the value of the maximised likelihood function and the goodness of fit tests. The data base used for each model is also explained in this part.

Chart 3. Used or selected variables classified by groups

Classes and Groups of the 82 variables tested	Groups of selected variables			
	Security		Speed	
Equations	**A**	**G**	**μ(V)**	**σ(V)**
Behaviour (4 elements)				
G 1. Speed	∇√	∇√		
G 2. Perceived risk (Expected maximum utility)			∇√	∇√
Traffic (7 elements)				
G 3. Traffic flow	∇√	∇√	∇√	
G 4. Rank flow rate				
Road-Dimensions in height and lenghtwise (32 elements)				
G 5. Approaching zone in length	√	√		
G 6. Approaching horizontal zone	∇√	∇√		∇√
G 7. Nature of the in length element	∇√	∇√		
G 8. Nature of the horizontal element	∇√			
G 9. Difficulty associated with the "in length" element		√		
G 10. Difficulty associated with the horizontal element	√	√		
Road-Dimensions in width and section length (13 elements)				
G 11. Hierarchical rank in national network	√	√	∇√	∇√
G 12. Rank of profile (# of lanes)		∇√		
G 13. Lanes length and maintained width	√	√		
Road-Surface indexes (26 elements)				
G 14. Smoothness, friction, tracks, surface coating loss	∇√ ®	√ ®	∇√	
G 15. Structural : various cracking	√ ®	√ ®		∇√
G 16. Material used for surface layer: concrete				
G 17. Bonding used for surface layer : bitumen				∇√
Total number of tested elements	80	80	80	80
Number of elements selected	32	34*	5	7
Reference elements selected (**)	(5)	(5)	(2)	(2)
Box-Cox transformations used	4	1	2	1
Auxiliary variables obtaining a β, associated with $X_k \geq 0$ variables and transformed by λ	4	14	0	1
Regression constants	1	2	1	1
Number of observations	2541	1225	17	17

The symbol √ means that the elements of this group have been selected and that the numerical results are shown in details in Appendix 2, in the case of frequency and severity equations, and in Chart 11, for the speed equations. If by any chance, the group encloses one or several variables of confidential results, the symbol ® is added : identifying these variables by name is impossible in Appendix 2. The symbol ∇ means that the results for some of the variables are explained in the text.

(*) 34 variables but 68 coefficients β.

(**) Variables for which the coefficients are confounded with the regression constants coefficients.

5.2. Chosen results : safety equations

The proposed extracts present four models: models of first-step *prob:1* and *sever:1*, that do not enclose behavioural variables (expected speed measurements), and models of *prob:2* and *sever:2* where these variables are considered, according to (6) and (7). We will make general statistical comments on the models and specific comments on 11 variables belonging to each of the 5 classes of chart 3. The reader interested in getting more information will find in Vernier (1999) a discussion about the *level of traffic* (Group 3), the *in length profile of the approach zone* (Group 5), the *heading angle* used as is or split up in geographical orientation (*north, south*, *east, west*; Group 10), as well as qualitative comments of the surface indexes such as the *gravel stone emergence*, the *localised* or *important repairs*, or the *slight* or *severe deformations* (Group 14).

A. Statistical comments

In a first instance, in Chart 4, we provide the general estimations statistics related to the number of observations and to the goodness of fit of these observations.

The probability model is based on a sample of 1 225 individuals that have been involved in an accident; they are also used for estimating the severity model. Furthermore, 1 316 accident free individuals are considered, completing at 2 541 observations the sample model of accident frequency. The structure of the endogenous stratification was taken by random drawing amongst the 1 316 individuals from the complete data base at our disposal. The weighting proportions used in the estimation are calculated on these basis. As for the adjustment, adding the variables of behaviour allows, as far as the χ^2 test is concerned, a more observable gain in the model related to severity compared to the model related to accident frequency.

Chart 4. Estimation of the frequency and severity: adjustment to the observations

	model prob:1	*Model prob:2*	*model grav:1*	*model grav:2*
III. General statistics				
Final Log-likelihood	-725.669	-724.371	-1064.15	-1061.46
Log-likelihood with constant only	-847.365	-847.365	-1195.99	-1195.99
Rho squared	.144	.145	.110	.112
Rho carré bar: Akaike	.099	.097	.043	.041
Horowitz	.121	.121	.076	.077
Hensher and Johnson	.131	.131	.080	.080
Good prediction percentage	55.057	55.293	57.143	57.796
Number of alternatives	2	2	3	3
Number of observations	2541	2541	1225	1225
Estimated parameters of - Variables	30	32	66	68
- Constants	1	1	2	2
-associated binary	4	4	14	14
-Box-Cox transformations	3	4	1	1

B. Behaviour: the expected speeds

We will find in Chart 5 the <u>*probability points*</u> calculated according to equation (23) and evaluated to the average sample of variables, as well as the Student's statistics of first group variables, which is the expectation of mean speed and of standard deviation of the speeds. To get a proper interpretation of a Student's *t*, we have to determine if the coefficient to which it refers is a structural coefficient or a difference of two coefficients. Indeed, in the severity equations, the Student's *t* statistics test the difference between the structural coefficients of the variables for the indicated condition and those of the reference condition indicated by *(w.r.t.)* for an inclusive variable (following the χ^2 tests performed). In the accident frequency equation, the reference condition is the non accident case. In every case, the Student's *t* does not indicate much compared to the exact information given by the χ^2 test and can therefore contradict it.

The estimated mean speeds and the estimated expected standard deviation of speeds explain both accident probability and severity, according to the χ^2 tests, even if the indication of Student's *t* added leads us to believe that the results are more significant in the case of speed than in the case of standard deviation.

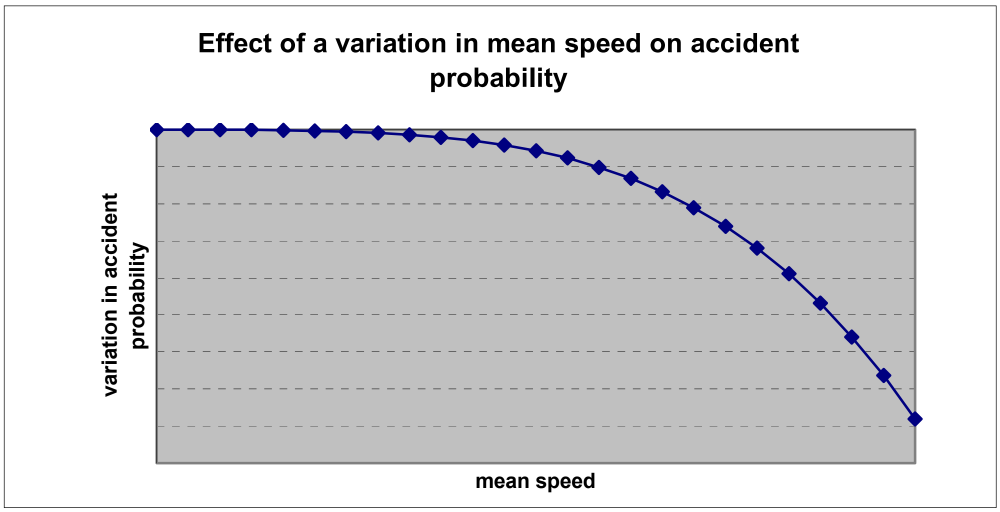
Chart 5. Behaviour effects: mean speed and standard deviation of the speed

I. **Accident probability *(a)***		**Model *prob:2***		
Probability points and Student's t of var.		*π(a)*	*t* = 0	*t* = 1
G 1. Speed				
expected mean speed	VM	-.202	(-1.63)	
expected standard deviation of speeds	ET	.001	(-.03)	
Box-Cox transformation of VM	λ_4	4.233	[1.73]	[1.32]

I. **Probability of severity *(g)*:** 1. *slight injuries = (B .L.)* 2. *severe injuries = (B.G.)* 3. *fatal = (M.)*		**Model *grav:2***			
Probability points and Student's t of var.			*π(s)*	*t* = 0	*t* = 1
G 1. Speed					
expected mean speed	VM	*(B .L.)*	-.247	*(w.r.t.)*	
		(B.G.)	.138	(1.37)	
		(M.)	.108	(1.54)	
expected standard deviation of speeds	ET	*(B .L.)*	.029	*(w.r.t.)*	
		(B.G.)	.023	(0.01)	
		(M.)	-.052	(-1.31)	
Box-Cox Transformation of VM	λ_4		1.802	[0.29]	[0.13]

Speed. In terms of probability, an increase in mean speed reduces significantly the accident probability, every other factors being equal. Moreover, the estimated function form relating to the effect of the mean speed on the accident probability differs statistically from a logarithmic form *(λ=0)* with a degree of confidence greater than 10%. In return, we do not set aside the linear form. The model however estimates a function to the fourth power, which would indicate that an increase in speeds has a small impact if the initial speed level is low and the effect will be greater at high speeds, as shown in Figure 5. This result is coherent with the anticipations, insofar as speed and attention are complementary and relatively few accidents occur at high speeds.

Figure 5. Effect of a variation in mean speed on accident probability



In terms of severity, we can notice from the model that the mean speeds play a role in increasing the conditional probability of severe or fatal accidents, even though the estimated coefficients for the differences between these conditions (severe or fatal) are not very significant. The model tells us that if the speed doubles, the accident severity also increases. It is explained by drops of approximately 25 points in the probability of having an accident with slight injuries, and lower by almost equal probabilities of having an accident with severe injuries or having an accident with fatal injuries.

Standard deviations of speeds. According to Student's test on the difference between the accident conditions (severe or fatal), the standard deviations of speeds have an effect on the "fatal accidents" significantly different from the one observed on "accidents with slight injuries". Increased standard deviations seem to reduce the probability of having a fatal accident, but increase the two concurrent probabilities. This situation can be explained by drivers paying more attention to the road when the traffic flow speeds are very different and by different accident typologies.

The importance of the standard deviation is much less than that of the mean speed, which is the most discriminating speed measurement for road safety. The results for these two behavioural parameters appear consistent with those of Solomon (1964) who finds a positive correlation between the accident frequency (without severity distinction) and the standard deviation of speeds.

C. Traffic: the percentage of heavy goods vehicles (HGV)

The second result we plan to analyse is related to the percentage of heavy goods vehicles in the traffic flow. We assume that the traffic composition (types and volume of vehicles) has an effect on road safety. In Chart 6, this assumption is confirmed : the variable giving the percentage of heavy goods vehicles presents a significant explanatory power for the frequency and severity of accident, even where Student's *t* tests are concerned.

Chart 6. Traffic effects: the percentage of heavy goods vehicles

I. **Accident probability *(a)***		**Model *prob:2***	
Probability points and Student's t of var.		*π(a)*	*(t* = 0)
G 3. Traffic flow (volume)			
percentage of heavy goods vehicles	PourcPL	-.110	(-3.66)

I. **Probability of severity *(g):*** 1. *Slight injuries = (B.L.)* 2. *Severe injuries = (B.G.)* 3. *Fatal = (M.)*		**Model *grav:2***	
Probability points and Student's t of var.		*π (g)*	*(t* = 0)
G 3. Traffic flow			
percentage of heavy goods vehicles	PourcPL	*(B .L.)* -.025 *(B.G.)* -.048 *(M.)* .073	*(w.r.t.)* (-0.33) (2.45)

Probability. Concerning the probability of having an accident, we observe that an increase in the percentage of heavy goods vehicles in traffic flow corresponds to drops in accident probability; if this percentage doubles, the possibility of having an accident loses eleven points.

In general, the increase in heavy traffic produces two adverse effects on the accident probability. Certain studies confirm a reduction in the accident probability. This phenomenon can be explained by the driving experience of HGV drivers and the special stability and horsepower characteristics of these vehicles. These two elements affect and change the driving practices (calmer, slower speeds, anticipation). In return, other studies show different results and support them by stating that the users take an extra risk when executing hazardous overtaking or when driving with more stress. The context described by the data base used, shows a reduction in accident probability when the percentage of HGV increases.

Severity. In terms of severity, we observe that the impact of the HGV percentage is not significantly different for the first two alternatives (slight and severe injuries). In return, the impact is very different on the probability of having a fatal accident. The higher the proportion of

HGV, the more the accident severity increases, in particular, the probability of having a fatal accident. But we have to single out two cases. When the percentage of HGV increases in traffic flow, two main accident types are possible. The first concerns accidents involving HGV and the second involves only light vehicles.

In the first case, it is clear that the severity of injuries will be more important. In fact, we often define a traffic collision as an exchange of energy between two vehicles. The greater the developed energy, the more severe the accident. Therefore, accidents involving HGV appear more serious. Besides, several of the HGV characteristics worsen this phenomenon. Although the mass differential between light and HGV play a great part in increasing the severity, the HGV structure is also a factor. We notice that the bumpers on HGV are still at the windscreen level of light vehicles. In case of an impact, sideways or head-on, occupants of light vehicles are poorly protected by body work which does not act as an energy “absorber”. This problem has been pointed out and HGV are now fitted with a lower bumpers. However, the overall vehicle fleet has not been renewed. In the second case, the severity of injuries is connected with the type of accidents that occur in that context, especially head-on impacts between vehicles due to risky overtaking manoeuvres.

D. Road-dimensions in height and longitudinals: turns and crown

Three-turn sequence ; crown of a variable radius. We will notice in Chart 7 that two variables, linked to height or longitudinal, reduce the accident probability but increase the accident severity, even after considering speed adjustments. It is an approaching zone constituted of a three-turn sequence (compared to a practically straight one), and of a “bumped” crown zone, or if you prefer of variable radius (compared to a climb one).

The horizontal elements : circle to the left. Other variables in length only influence the accident frequency. We, in fact, studied the impact of different geometric horizontal elements, since readings from the gyroscopic stations allow us to observe 5 horizontal elements: straight alignment, circle to the right, circle to the left, connections to the right and connections to the left. We have set up in Chart 8 the traffic risks related to the last four elements mentioned, in reference to a movement on a straight section.

First of all, the results presented in Chart 8 show that the element on which road users travel has an impact on accident probability. Travelling on a connection has a very different effect compared to travelling on a straight section; travelling on a turn has no significant impact. We can see from the model that for traffic entering or exiting (from turns) the accident probability is reduced, compared with travelling on a straight section.

Chart 7. Road effects: three-turn sequence and crown of a variable radius

I. **Accident probability *(a)***		**Model *prob:2***	
Probability and Student's t of var.		*π(a)*	*t* = 0
G 6. Approaching horizontal zone (w.r.t. : practically straight)			
Three-turn sequence	Virvirvir	-.082	(-1.77)
G 7. Nature of the in length element (w.r.t. : climb)			
Bumped connection	Racbos	-.144	(-2.12)

I. **Probability of severity *(g)*:** 1. *Slight Injuries* = *(S .I.)* 2. *Severe Injuries* = *(S.I.)* 3. *Fatal* = *(F.)*		**Model *grav:2***	
Probability points and Student's t of var.		*π (g)*	*t* = 0
G 6. Approaching horizontal zone (w.r.t. : practically straight)			
Three-turn sequence	Virvirvir	*(B .L.)* -.152 *(B.G.)* .113 *(M.)* .040	*(w.r.t.)* (1.75) (1.39)
G 7. Nature of the in length element (w.r.t. : climb)			
Bumped connection	Racbos	*(B .L.)* -.181 *(B.G.)* .096 *(M.)* -.085	*(w.r.t.)* (1.15) (1.94)

This result is subject to a number of reservations namely where accident localisation is concerned. When a vehicle loses control and runs off the road, at least two accident localisations are possible: the place where the driver loses control and the place where the vehicle runs off the road. These two possibilities can be explained by a location on a turn or on a connection. It is hard, sometimes risky, to isolate these two parameters.

Chart 8. Road effects: the horizontal geometric elements

I. **Accident probability *(a)***		**Model *prob:2***	
Probability points and Student's t of var.		*π (a)*	*(t* = 0)
G 8. Nature of the horizontal element (w.r.t.) : straight section)			
Circle to the right	cerdro	.004	(0.12)
Circle to the left	cergau	.031	(1.05)
Connection to the right	racdro	-.168	(-3.60)
Connection to the left	racgau	-.068	(-1.75)

However, a closer analysis of the data base allowed us to clearly identify an important confusion factor. We can see that connections indicate the presence of short radius circles, while the circles represent long radius turns that do not need connections. The model therefore stipulates that long radius circles do not imply more risk than straight sections. This result is in accordance with our expectations. On the other hand, the estimates indicate that connections are safer than circulation on straight section. This result first appears surprising but the connections are associated with the presence of clothoids in approaching the turn (the road users are well informed on the difficulties associated with the curve section) and in exiting the turn (the road users can slowly go back to their normal speed). These elements do not surprise the user and the road appears safe for negotiating with the geometric elements.

In a second phase, the model demonstrates that the direction of the element also plays an important part. We observe, in fact, that curves to the left are more dangerous than curves to the right.

Connections to the right are safer (-16,8 % of accident probability) than connections to the left (- 6.8 %). A similar result has already been presented in a SETRA-CETE Normandie Centre document (1997). The authors explain this situation by the fact that in curves to the left, lateral acceleration is applied from the right and is amplified by the cross-fall, generally sloping to the right. If the driver loses control, his vehicle goes in direction of the right-hand shoulder where the recovery manoeuvres are limited and depend on the characteristics of the shoulder in terms of surface and the presence of obstacles. In curves to the right, lateral acceleration is applied towards the left and, in most cases, the cross-fall slows down the effects of this acceleration. In case of control loss, the vehicle slews onto the oncoming traffic lane. If there is no traffic, the carriageway becomes a wide recovery zone, surfaced and without any obstacles.

The results obtained confirm that there are more accidents on curves to the left than on curves to the right. We have to consider that, as indicated in Chart 3, these variables from Group 8 had no effect on accident severity.

E. Road-dimensions in width: three lanes

However, other variables influence the severity but not the frequency. Three-lane roads (compared to two-lane roads) increase accident probability with severe injuries but reduce the severity of other accidents, even if there is no real significant effects : we have to assume that there is a risk compensating factor which explains the sign pattern in Chart 9.

Chart 9. Road effects: three-lane road

I. **Probability of severity *(g)*:** 1. *Slight Injurie s* = *(B.I.)* 2. *Severe Injuries* = *(B.G.)* 3. *Fatal* = *(M.)*		**Model *grav:2***	
Probability points and Student's t of var.		π *(g)*	(*t* = 0)
G 12. Profile class (w.r.t. : two lanes)			
Three lanes	troisv	*(B .L.)* -.031 *(B.G.)* .040 *(M.)* -.010	*(w.r.t.)* (0.67) (-.10)

F. Surface Road-index: smoothness

Amongst the surface indexes, we will notice, in Appendix 2 that the *smoothness* variable, which describes the pavement smoothness, affects only the accident probability. What makes it so interesting is its close relationship with travel speed : the higher the smoothness, the more the pavement becomes degraded by longitudinal irregularities and driving at high speed becomes difficult. It is not surprising to find in Chart 10 that taking into account speed lowers considerably the value of the smoothness variable on the two concurrent probabilities.

Chart 10. Road effects: smoothness

G 14. Smoothness, friction, tracks, surface coating loss...				
I. **Accident probability *(a)***		**Model *prob:1***		
Probability points and Student's t of var.		*π (a)*	*t* = 0	*t* = 1
Index of pavement planeness	valuni	-.070	(-1.42)	
associated binary	bvaluni	.074	(1.23)	
B-C Transformation of valuni	λ_1	.535	[2.77]	[-2.40]
I. **Accident probability *(a)***		**Model *prob:2***		
Probability points and Student's t of var.		*π (a)*	*t* = 0	*t* = 1
smoothness	valuni	-.035	(-.60)	
associated binary	bvaluni	.073	(1.12)	
B-C transformation of valuni	$?_{?1}$	.525	[2.56]	[-2.31]

If we observe, in the first model, that the poorer the smoothness, the lesser the accident probability (the value of 1,42 render the importance of the effects irrelevant but recognises the direction), we can not find a similar result in the second model. Consequently, it is obvious that in the first model the non controlled effect of the vehicles speed is enclosed in the smoothness variable. This variable estimate is therefore biased. Besides, the maximum likelihood function of model *prob:2* can be compared to the one of the identical model in which we have left out the smoothness variable. The difference between these two likelihood functions equals 0,62, which means that the smoothness variable plays no role in the accident probability. The only variable that is not affected by adding the speed parameters is the Box-Cox transformation parameter to which the variable is subjected. However, this parameter is also applied to the variable of the current section length, belonging to Group 13, where the estimation is not affected by speed consideration. Because of the estimation of this parameter does not only rely on the smoothness variable, the value of the parameter can vary from one model to the other.

We observe that a purely mechanical model linking the pavement characteristics to the accident probability gives a value to the smoothness variable that actually includes the speed effects. Consequently, smoothness affects the road user's perception and forces him to adjust his speed according to the driving conditions available, but has *no real direct impact* on accident probability.

G. Surface or structure road-indexes

We can see in Appendix 2 that all the surface or structure variables have an associated binary variable. These variables are equal to one when the continuous variable is positive and equal to zero otherwise. In the case where the continuous variable is shown in linear form, the associated binary variable represents the effects of the presence of the phenomenon and the continuous variable captures the effect of the level of the phenomenon. Nevertheless, when the continuous variable is subjected to a Box-Cox transformation, the binary variable assures the invariance of the transformation to the unit measurements used to describe the phenomenon. The reader will notice that the binary variables signs are generally opposed to the ones of their continuous variables, with the exception of the variable G (partially) and the variable I (totally). Interpreting the effects of these variables requires that we combine the effects of the sets of variables.

5.3. Complete results: speed equation

Modelling of the mean speed measurements, *VM*, and the standard deviation of speeds, *ET*, allows us to link them to perceived risk and to infrastructure variables associated with driving comfort. Based on the analysis of only 17 present sections with different characteristics, our *estimations are to be considered with circumspection*. The two models estimated displayed in Chart 11 generate, in accordance with (8) and (11), expected speed measurements for sections from the estimation sample of equations (6)-(7).

The first part of this chart presents the *elasticity,* computed with respect to equation (24) and evaluated to the average of the sample, and the Student's *t* associated with the estimators selected by the tests of the χ^2 . The second part specifies the functional forms used and two tests of hypothesis relating to linearity and logarithmic form. Finally, the third part shows results related to the goodness of fit.

A. Statistical comments

We first notice that the Student's *t* of the Box-Cox transformation, calculated with the Berndt *et al*. (1974) first derivatives method, seem roughly estimated here - probably because of the low number of observations - they do not represent a clear gain in log-likelihood : we would rather trust the χ^2 tests that are exact, and suggest a logarithmic form of the transformed terms. The "smoothness", "severe transversal cracks" and "perceived risk of severity" variables give a low gain in likelihood when they are not transformed, in either equations.

However, the Box-Cox transformation, associated with the two inclusive values of frequency and severity in the mean speed equation, is near 0 : we can therefore accept them as expected utility measurements, defined in the strict sense by the natural logarithmic of the inclusive value. Yet, in the standard deviation equation, the linear form of the inclusive values is selected, but these measurements are less significant.

Chart 11. Result of speed equation estimation

		Model VM		Model ET	
I. Elasticity and Student's t of variables		***η***	***(t = 0)***	***η***	***(t = 0)***
G 2. Perceived risk					
Perceived probability risk	EMUp	-.207	(-.68)	.047	(.07)
Box-Cox transformation of EMUp	LAM2	[λ_2]			
Perceived severity risk	EMUgr	-.124	(-1.67)	.043	(.52)
Box-Cox transformation of EMUgr	LAM2	[λ_2]			
G 3. Traffic volume					
aver. annual daily traffic all veh.	MJA	-.071	(-.56)		
G 6. Approaching horizontal zone (w.r.t. : straight section)					
turn	virage			-.226	(-1.76)
G 11. Hierarchical classification within the (w.r.t. : national road)					
Superhighway stretch	LACRA	.243	(1.39)	.450	(3.02)
G 14. Smoothness, friction, tracks, surface coating loss...					
Index of pavement planeness	valuni	.242	(1.50)		
Box-Cox transformation of valuni	LAM1	[λ_1]			
G 15. Structural : various cracking					
Severe transversal cracks	ftgrav			-.002	(-2.01)
Box-Cox transformation of ftgrav	LAM1			[λ_1]	
associated binary variable	dftgrav			.535	(6.85)
G 17. Bonding used for surface layer : bitumen,...(w.r.t. : bitumen-bound graded aggregate)					
Cement-bond aggregate	GH			.230	(0.76)
Untreated small gravel	NT			-.248	(-1.59)

II. Values of λ□; Student's t	[λ]	**t = 0 ; t = 1**	[λ]	**t = 0; t = 1**
Transformation λ_1	-.135	-.04 ; -.34	-9.99	-.01 ; -.01
Transformation λ_2	-.327	-.27 ; -1.08		

III. General statistics	Model VM	Model ET
Log-likelihood with λ estimated	**-57.185**	**-37.431**
Log-likelihood with λ = 1	-59.842	-45.864
Log-likelihood with λ = 1 and variable **dftgrav** included		-42.856
Pseudo-R-2	.746	.861
Pseudo-R-2 adjusted for degrees of freedom	.549	.683
Number of observations	17	17
Number of estimated parameters		
- Betas - Variables	5	7
- constants	1	1
- associated binary variables	0	1
- Box-Cox Transformations	2	1

A third thing that can be observed is that all infrastructure variables selected by the χ^2 tests are associated with the driving comfort : in presence of perceived risk variables, the tests have indeed excluded the other infrastructure variables, as anticipated in Figure 2 by the expression (k(N-I)).

Fourthly, concerning results relating to the estimation of the standard deviation model, we observe that the binary variable " dftgrav ", associated with the " ftgrav " variable, generated by the presence of Box Cox transformation on " ftgrav ", is particularly significant. So, we tested the impact of this variable introduced directly in the linear form of the model : in that case, the binary variable " dftgrav " represents only the *presence* of severe transversal cracks, and the variable " ftgrav " represents the *level* of the cracks: the value of the likelihood function obtained (- 42.856) confirms the role of the presence of severe transversal cracks, as well as the role of the level of cracks, since gain in log-likelihood is of 3 points (compared to - 45.864).

B. The mean speed equation

Parameters affecting mean vehicle speed measured at a point are the perceived risk variable, the global vehicle traffic expressed as the average daily number of vehicles, the type of lane and the smoothness. As expected, the infrastructure characteristics, other than those linked to comfort, have not been selected by the χ^2 tests.

The perceived risk measures. The concept of perceived risk is an important factor linked to the accident phenomenon. The perceived risk shows the extent (width), evaluated by the driver, of the difficulties generated by infrastructure. Therefore, we assume that when the driver senses a greater risk, he is more attentive and we can expect him to reduce his travel speed.

The estimated coefficient relevant to the perceived risk in terms of accident probability has no significant effect different from zero on the mean speed, unlike the perceived risk in terms of severity. Still, as expected, the model shows that with a rise in severity perceived risk comes a drop in mean speed.

We discover that the road users react strongly to an additional risk. A 10 % increase of the level of risk, concerning the potential severity of the accident, produces a 12.4 % reduction in speed. This result is evaluated with other factors being equal.

The important effect of the perceived risk linked to severity indicates that the motorist has a certain comprehension of danger. For the driver, the mean speed seems to be more associated with a level of damage. We often put forward the example of highways in which the road users feel that accidents are infrequent, but when they occur, they are far more serious and even have fatal consequences. From these findings we can presume that the drivers, aware of the eventual severity when an accident occurs, are more careful and there are less severe injuries and fatalities. This behaviour results in a high reduction in speed and by a stronger reaction to a severity level than to an accident probability.

Moreover, the model shows that the estimated form of the mean speed adjustments as a function of risk perception is close to the inverse square root function. Even though the *t* hypothesis tests, for which we earlier said that they were mathematically weak, do not allow us to throw out neither the linear or logarithmic form, the χ^2 test, always exact, shows that the Box-Cox transformation offers a significant gain in likelihood. If we accept the estimated form according to a power of - *0.33*, which indicates that above a certain risk level, the impact of a speed change tends to stabilise while at low risk levels, the same change has a big impact.

This behaviour is explained by several elements. First, we can suppose that a risk habituation phenomenon occurs. In a safe environment without any constraints, the emergence of a lesser hazardous factor will have a major effect. *A contrario*, in a very dense and complex road environment, this additional risk will not have any comparable effect. This situation is compounded by the legal speed limit levels. We suppose that motorists feel safe driving at a speed close to the legal limit and do not want to drive slower. The main explication is that speed is not only justified by infrastructure parameters but also by characteristics concerning drivers and their vehicles.

Traffic. The sections under study are located in rural areas. Traffic is free-flowing and there are no traffic build-ups. In that context, we are particularly interested in the way the network average user behaves when faced with different traffic densities. In all probability, we can expect that a greater number of vehicles will oblige the driver to reduce speed. The estimation does not enable us to accept this anticipation, because of the lack of significance of *t* associated with the estimated coefficient of this variable though selected in the list.

Smoothness. The smoothness qualifies the **planeness** level of the pavement. A good level of smoothness characterises a perfect pavement on which there is no discomfort felt. The estimated parameter presents a low level of signification (1.50) which indicates that the variance associated with the estimation is relatively important. Despite this uncertainty, the estimation is as we expected. The model reveals that a smoothness degradation results in a reduction in speed. Lots of studies confirm this result (Gambard, 1986). Smoothness defects produce vibrations in the vehicle bodywork that are perceptible by the driver. For comfort reasons, the driver slows down spontaneously.

C. The equation of standard deviation of speeds

The standard deviation depends mostly on the geometric road characteristics (turn, number of lanes), on the presence or level of severe pavement deformation, and on the type of pavement (untreated, bitumen-bound graded aggregate, cement-bond aggregate).

The perceived risk. Unlike the mean speed, the standard deviation of speeds is not significantly influenced by perceived risk measurements in probability or in severity. This fact can be explained by the different driving abilities of the road users or the vehicles they use. Most motorists will reduce their travel speed when they feel there is a risk. Yet, speed is not reduced with the same intensity by everybody and it does not really affect the standard deviations.

Vehicle description (safety equipment, horsepower, age, type…) and the driver characterisation (age, sex, reason of the trip…) are the two groups of variables that would allow a more in-depth analysis. Unfortunately, the data base at our disposal does not contain this information and does not enable us to establish the responsibilities linked to these parameters in the speed differentials.

Carriageway type. The carriageway type contains several road characteristics. It refers to the width and the number of lanes, to the level of traffic, and to the infrastructure safety equipment. The 17 sections used cannot be significantly represented in the five existing classes, found in Appendix 1 : Liaison Autoroutière en Continuité du Réseau Autoroutier (LACRA; links ensuring completion of the motorway network) Grande Liaison d'Aménagement du Territoire (GLAT; land development motorway links), Voie Routière Urbaine (VRU; urban express roads), Route Nationale de Liaison (RNL; trunk roads), and Route Nationale (RN; national roads).

We have grouped different road classes. The first one contains RN and RNL, the second one contains the other types of roads. We may have reservations about the urban express roads, but since this type of road is usually a dual two-lane carriageway we have chosen to include it in the above category. However, the possible prediction errors will have little impact on the final results of the study owing to the low number of these sections in the data base.

The carriageway type has a major impact on speed deviations. On dual two-lane sections, the model presents mean standard deviations in speed superior than 45% compared to standard errors on roads with only two lanes. This result matches our expectations. The major difference is explained by the differences in horsepower and comfort of the vehicles on the market and by the different drivers; the most important difference is the one between light and heavyweight vehicles. Still, in a same range of vehicles, considerable differences exist in terms of horsepower, equipment and comfort. As in the risk variable analysis, the lack of information on the driver and his vehicle does not help explain this result which could also be modulated if these variables were included in the model.

The bonding nature and severe transversal cracks. Transversal cracks and the bonding nature are two pavement structure characteristics. Bonding is the material used to bind the gravel forming the pavement inferior coating. There are three types of bonding: Untreated (NT), Bitumen-bound graded aggregate (GB) and cement-bound aggregate (GH). The untreated pavements (NT) often have a high level of deformation, especially when traffic is heavy.

The NT pavements are more subject to cracking than GB or GH structures. This is due to the basic layer composition where the gravel is not heavily fixed. Moreover, there is, between GB and GH structures, considerable differences, also linked to the cracking phenomenon. The GH structures are solid and subject to expansion due essentially to temperature changes. This phenomenon is characterised by the emergence of transversal cracks in the upper surface layer. In return, the GB pavements, more flexible, are not affected by the climatic changes and are less exposed to cracking.

The model shows that the standard deviations of speeds differ according to bonding natures. More precisely, driving on a pavement with a GB or GH bonding has no significant impact. However, standard deviations on untreated pavements are significantly lower than those observed on GB pavement. This result is due to the cracks appearing on pavements and can be confirmed by the results concerning the severe transversal cracks parameter.

These variable impacts are of two kinds: the level and the presence of cracks. The model reveals that it is more the presence of severe cracks than their level which has a major impact. This is confirmed by the binary variable associated with the parameter characterising the presence or absence of cracks on the section.

Cracks on pavement increase the standard deviation of speeds by more than 50 %. It appears that motorists do not adjust their speed with the same sensitivity to this pavement surface characteristic. The transversal cracks are visible by the users. It is possible that road users associate with them an insufficient pavement safety level.

The two results concerning the pavement structure are therefore consistent.

D. Conclusion

All the results met our expectations and, even though the estimation is based only on 17 sections, the mean speed values and standard deviation values generated are not outliers. The inferred speeds may be imperfect because certain sections of the sample show particularly severe geometrical configurations or influential surface characteristics not included in the model. Our data base is certainly a limiting factor for a more accurate inference but no absurd value is generated and the associated speed histograms are very coherent.

6. General conclusion

We estimated a *simultaneous* system of behaviour interactions between chosen speed and perceived risk, by measuring it as a decomposable *expected maximum insecurity (EMI)*, according to Knight, between *uncertainty* and calculated *risk*. We have not found other road studies using any of these three dimensions.

Our results are generally strong because the *relevance* of the different factors has been established with the help of the χ^2 tests and their inclusion *form* determined as endogenous by the Box-Cox transformations. In particular, the logarithmic form, called *log-sum*, required by the expected maximum utility theory has proven optimal in representing the perceived risk in the mean speed equation. The Box-Cox transformations were again useful, for their goodness of fit to the data as well as for their more reasonable results obtained, in comparison to results of predetermined form, particularly linear.

The exceptional richness of the data base in descriptive variables of geometry and road conditions enabled a first sorting out of the problem, in a multi-level context of DRAG type. Nevertheless, the absence of variables on other factors such as the driver, the vehicle and the climate characteristics, as well as the poor representation of the speed sample, undoubtedly imply *biased results.*

We propose a theoretical structure to establish a very complete behaviour model, thanks to the stratification made on the sample in Logit model. This stratification enables us to complete the data base without generating exorbitant cost, on a financial level as well as a human level. A survey on 1 500 road users allows, in a first instance, to take into consideration the characteristics linked to drivers and their vehicles. To be able to use the estimated simultaneous structure, an additional survey on users' travelling speeds would be necessary. Although it is less costly than an overall survey, a proper collection over 1 500 road sections would generate important expenditures. We could therefore obtain a model with credible and *usable* results, instead of only being interesting.

With the addition of new explanatory variables, the model would particularly help us determine the *respective uncertainty and calculated risk parts* explaining the observed risk. This clear distinction opens the way for intervention on the part of road safety authorities such as signage, warnings, as well as the determination of the limits of these interventions. If certain corrective measures can in theory reduce uncertainty, there is a limit to this reduction, or an irreducible "fond d'incertitude": our structure would isolate it.

The advantage of interventions of type " uncertainty reduction " is no less important than the corrective measures directly linked to travel speeds, concerning the consequences on probability and severity of accidents.

7. References

ALJANI, A. A. M., RHODES, A. H. and A. V. METCALFE, " Speed, Speed Limits and Road Traffic Accidents under Free Flow Conditions ", Accident Analysis and Prevention, 31, 161-168, 1998.

AMEMIYA, T., *"Advanced Econometrics"*, Harvard University Press, Cambridge, Massachusetts. 1985.

ARELLANO, M. and C. MEGHIR, *" Female Labour Supply and On-The-Job Search : An Empirical Model Estimated Using Complementary Data Sets "*, Review of Economic Studies, 59, 537-557, 1992.

ATO EGUAKUM, G. and F. WILSON, *"Modelling Risk as a Choice Process"*, Compte rendu de la 9[eme] conf. canadienne multidisciplinaire sur la sécurité routière, pp. 173-184, 23-25 mai 1995, Montréal. 1995.

BADEAU, N., BAASS, K. et C. POIRIER, *«Stabilité des véhicules dans les courbes: Expérimentations, analyse et interprétation des résultats»*, Non Publié, Centre de Recherche sur les Transports, Université de Montréal, 75 p., janvier 1997.

BARBER, P., BADEAU, N. et K. BAASS, *«L'effet des profils horizontal et vertical sur la sécurité routière, Tome I »,* Non Publié, Laboratoire de circulation et de sécurité routière, École Polytechnique*,* Montréal, 73 p., décembre 1998.

BEN-AKIVA, M. and S. R. LERMAN, *"Discrete Choice Analysis: Theory and Application to Travel Demand Analysis"*, Massachusetts, MIT Press. 1985.

BERGERON, R., *Le Livre Noir de l'Automobile*, Hypothèse, Montréal, 435 p., 1999.

BERNDT, E. R., HALL, B. H., HALL, R. E. and J. A. HAUSMAN*, " Estimation and Inference in Nonlinear Models "*, Annals of Economic and Social Measurement, 3, 653-665, 1974.

BOLDUC, D. and S. BONIN*, " Bayesian Analysis of Road Accidents: A General Framework for the Multinomial Case "*, in Dionne, G. and C. Laberge-Nadeau, eds, Automobile Insurance: Road Safety, New Drivers, Risks, Insurance Fraud and Regulation, Kluwer Academic Publishers, Boston, MA, pp. 79-100, 1999.

BOYER, M., DIONNE, G. and C. VANASSE, *"Econometrics Models of Accident Distribution"* in G. Dionne (ed.) Contributions to Insurance Economics, Kluwer Academic Publishers, Boston, MA, pp.169-213, 1992.

BOX, G. P. and D. R. COX, *"An Analysis of Transformations"*, Journal of the Royal Statistical Society, Series B, 26, 211-243. 1964.

BUTT, A. A., SHASHIN, M. Y., CARPENTER, S. H. and J. V. CARNAHAN, *"Application of Markov Chain Process to Pavement Management Systems at Network Level",* Third International Conference on Managing Pavement, Transportation Research Board, Vol. 2, pp.159-172. 1994.

CARDOSO, J. L., *"Accident rates and Speed Consistency on Horizontal Curves in Carriageway Rural Roads"*, International Conference Traffic Safety on two Continents, 22-24 Dec. 1997, Lisboa. 1997.

COLLINS, K. M. and R. A. KRAMMES, *"Preliminary Validation of a Speed-profile Model for Design Consistency Evaluation"*, TRB, Preprint 960590. 1996.

CONTI, T. C., McMINIMEE, J. C. and P. SENEVIATRE, *"Maintenance Skid Correction Program in Utah",* Transportation Research Record 1410, pp. 32-40. 1990.

DELHOMME , P., et G. MALATERRE, *«Analyse bibliographique des méthodes d'investigation du comportement de conducteur par observation embarquées»*, Programme DRIVE V1040, INRETS, Arcueil. 1990.

GAMBARD, J. M., *«Approche du comportement cinématique du véhicule libre sur un axe routier»*, Thèse de docteur ingénieur, E.N.P.C., 1986.

GAUDRY, M., *«DRAG, un modèle de la Demande Routière, des Accidents et de leur Gravité, appliqué au Québec de 1956 à 1982»,* Publication CRT 359, Centre de recherche sur les transports, Université de Montréal, 1984.

GAUDRY, M., *"Some Perspective on the DRAG Approach and Family of National Road Safety Models",* Working Paper 9808, Bureau d'économie théorique et appliquée, Université Louis Pasteur, et Publication CRT 9807, Centre de recherche sur les transports, Université de Montréal. 1998.

GAUDRY, M. *et al*., *"TRIO Tutorial, Reference and Model Types Manuals, Version 2.0"*, , Publications CRT 902, 903, 904, Centre de recherche sur les transports, Université de Montréal, 1995.

GODLEWSKI, D., *«Optimisation de la gestion routière: Utilisation de l'uni longitudinal»,* Thèse de docteur ingénieur, E.N.P.C., 1985.

JAEGER, L., *«Développement d'un modèle explicatif des accidents de la route en France sur une base mensuelle de 1956 à 1993»*, Thèse de doctorat, Université Louis Pasteur, Strasbourg. 1998. Publication CRT-99-11, Centre de recherche sur les transports, Université de Montréal, 1999.

KANG, K., *" Ordered Probit Model of the Speed Selection Behavior : Results Based on a Korean Micro Data ",* Fourth Annual Conference on Transportation, Traffic Safety and Health, Tokyo, Japan, October 1998.

KNIGHT, F. H., *Risk, Uncertainty and Profit*, Houghton, Mifflin, 1921.

KULKARNI, R. B., *"Dynamic Decision Model for Pavement Management System",* Transportation Research Record 997, pp.19-27. 1984.

KULKARNI, R. B., GOLABI, K., FINN, F. and E. ALVITI, *"Development of a Network Optimization System"*, Arizona Department of Transportation, Final Report, Vol. 1 et 2, Woodward-cycle Consultants, San Francisco. 1990.

LAMM, R. and E. M. CHOUEIRI, *"Rural Roads Speed Inconsistency Design Method. Part 1. Operating Speed and Accident Rates on Two-Lane Rural Highway Curved Sections: Investigation about*

Consistency and Inconsistency in Horizontal Aligment", Clarkson University, Report 87026, Postdam. 1987.

LAMM, R., CHOUEIRI, E. M. and T. MAILAENDER, *"Comparison of Operating Speeds on dry and wet Pavements of two Lane Rural Highways"*, Transportation Research Record, 1280, pp.199-207. 1990.

LIEM, T. and M. GAUDRY, *"PROBABILITY: The P-2 Program for the Standard and Generalised Box Cox Logit Models with Disagreggate Data",* Publication CRT 527, Centre de recherche sur les transports, Université de Montréal, 1993.

MANSKI, C. F. and S. R. LERMAN, *"The Estimation of Choice Probabilities from Choice Based Samples"*, Econometrica, vol. 45. 1977.

McCARTHY, P. S., and S. MADANAT *"Highway Accident Data Analysis: Alternative Econometric Method"*,Transportation Research Record 1467, pp. 44-49. 1994.

OCDE, *«Systèmes de gestion des chaussées»,* Recherche en matière de route et de transports routiers. 1987.

OCDE, *«Recherche en matière de route et de transports routiers: Perspectives 2000»,* Recherche en matière de route et de transports routiers, Trentième anniversaire, 1996.

O'DONNEL, C. J. and D. H. CONNOR, *"Predicting The Severity of Motor Vehicle Accident Injuries Using Models of Ordered Multiple Choice",* Accident Analysis and Prevention, Vol. 28, pp. 739-753, 1996.

ORTUZAR, J. D. and L. G. WILLUMSEN, *Modelling Transport*, John Wiley and Sons. 1990.

PLANCO CONSULTING, *«Evaluation macro-économique des investissements dans les infrastructures de transport»,* Publications du ministère des transports, Cahier 72, Essen, Bonn. 1993.

SCHLESSELMAN, J., *" Power Families : A Note on the Box and Cox Transformation "*, Journal of the Royal Statistical Society, Series B, 33, 307-311, 1971.

SETRA-CETE Normandie Centre: *«Les accidents par sortie de chaussée»*, Document interne SETRA, 1997.

SETRA-CETE Méditerranée: *«Simulation de stratégies d'entretien sur le réseau national»*, Document interne SETRA, 1997.

SHANKAR, V., MANNERING, F. and W. BARFIELD, *"Effect of Roadway Geometrics and Environmental factors on Rural Freeway Accident Frequency",* Accident Analysis and Prevention, Vol. 27, pp. 371-389. 1995.

SHANKAR, V., MANNERING, F. and W. BARFIELD, *"Statistical Analysis of Accident Severity on Rural Freeways"*, Transportation Research Board, Preprint 960112, 1996.

SOLOMON, D., "Accidents on Main Rural Highways Related to Speed, Driver and Vehicle ", U.S. Department of Transportation, Federal highway Administration, July 1964.

UNDERWOOD, G. C., JIANG, C. and C. I. HOWARTH, *"Modelling of Safety Measure Effects and Risk Compensation"*, Accident Analysis and Prevention, Vol. 25, pp. 277-288. 1993.

VERNIER, K., " Fonctions de survie et méthode des M-estimateurs : application au module STRATÈGE du système français de gestion des chaussées GiRR ", Publication CRT-98-58, Centre de recherche sur les transports, Université de Montréal, novembre 1998.

VERNIER, K., " Les incidences du tracé routier et des états de chaussée sur le comportement des usagers vis-à-vis de la vitesse et de la sécurité ", Thèse de doctorat, Université Louis Pasteur, Strasbourg. 1999, et Publication CRT-99-16, Centre de recherche sur les transports, Université de Montréal, 1999.

WATANATADA, T., DHARESHWAR, A. M. and P. R. S. REZENDE-LIMA, *"Models for Predicting Vehicle Speed and Operating Costs based on Mechanistics Principals: Theory and Quantification",* HDM III, Volume II. 1987.

WEISS, A. A., *"The Effects of Helmet Use on the Severity of Head Injuries Motorcycle Accidents"*, Journal of The American Statistical Association, 87, pp. 48-56. 1992.

WILLIAMS, H. C. W. L., *" On the formation of travel demand models and economic evaluation measures of user benefit ".* Environment and Planning, 9a(3), 285-344, 1977.

ZEEGER, C.V. *et al.*, " Cost-Effectiveness Geometric Improvements for Safety Upgrading of Horizontal Curves ", FHWA-RD-90-021, Federal Highway Administration, McLean, Virginia, 237 p., 1991.

Appendix 1. Description of the 78 available variables

TRAFFIC VARIABLES

G. 3. Traffic volume (annual average) observed on the section	
pourcPL	Percentage of heavy goods vehicles
MJA	Annual daily average of traffic all vehicles
MJAPL	Daily average of traffic heavy goods vehicles

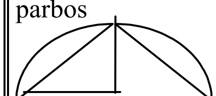
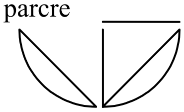
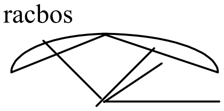
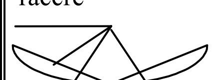
G. 4. Traffic level : allows to associate the section with a traffic class determined by an interval of annual daily averages.	
T0	MJA ∈ [750;2000]
T1	MJA ∈ [300;750]
T2	MJA ∈ [150;300]
T3	MJA < 150

INFRASTRUCTURE VARIABLES

G. 5. Approaching zone in length : describes the sections adjacent to the current section according to the in length profile. We have selected the adjacent sections on a length of 450 m.	
descente	Non-stop descent on 450 metres
montée	Non-stop climb on 450 metres
profalt	Altered profile, sequence of climbs and descents
lginc	Unknown profile of adjacent sections

[•••]

G. 6. Approaching horizontal zone : describes the adjacent sections according to the horizontal layout. We have selected the adjacent sections on a length of 450 m. The practically straight ones are circles with radius of more than 1000 m. Turns describe arches with radius of less than 1000 m.	
Qdqdv	Two practically straight alignments followed by a turn
Qdvv	Practically straight alignment followed by two turns
Virvirqd	Two turns followed by a practically straight alignment
Vqdv	Turn, practically straight alignment and turn sequence
virvirvir	Three-turn sequence
qd	Practically straight alignment

G. 7. Nature of the in length element : the in length element of the current section.	
dpente	descent
drampe	climb
parbos 	Section of constant radius making the transition between a climb and a descent.
parcre 	Section of constant radius making the transition between a descent and a climb.
racbos 	Section of progressive radius making the transition between a climb and a descent.
raccre 	Section of progressive radius making the transition between a descent and a climb.

[•••]

Appendix 1. Description of the 78 available variables (part 2)

[•••]

G. 8. Nature of the horizontal element: the horizontal element of the current section.	
cerdro	Circle (constant radius) turning to the right
cergau	Circle (constant radius) turning to the left
droite	Straight section
racdro	Exit or entry from a turn (progressive radius) turning to the right.
racgau	Exit or entry from a turn (progressive radius) turning to the left.
G. 9. Difficulty associated with the in length element.	
pen_ramp	Degree of slope (%)
G. 10. Difficulty associated with the horizontal element	
ray_vir	Turning radius (metres)
anglecap	Orientation according to geographic North evaluated in grads (Gr)
Est	Orientation to the East (50 to 150 Gr)
Nord	Orientation to the North (350 to 400 or 0 to 50 Gr)
Sud	Orientation to the South (150 to 250 Gr)
Ouest	Orientation to the West (250 to 350 Gr)
Nord-Est	Orientation to the North-East (0 to 100 Gr)
Nord-Ouest	Orientation to the North-West (300 to 400 Gr)

[•••]

[•••]

Sud-Est	Orientation to the South-East (100 to 200 Gr)
Sud-Ouest	Orientation to the South-West (200 to 300 Gr)

G. 11. Road type : indicates the group to which belongs the section. Classification made according to traffic, number of lanes and role of the section in the infrastructure network.	
GLAT	Land development motorway links
LACRA	Links ensuring completion of the motorway network
RN	National roads
RNL	Trunk roads
VRU	Urban express roads
G. 12. Profile class : gives the number of lanes in the current section.	
deux_2v	Dual two-lane section with divided pavements.
quatrev	Dual two-lane section with undivided pavements.
troisv	Three-lane section
deuxv	Two-lane sections
G. 13. Road variables : general elements related to the current section.	
long	Length of the current section
idroute	Number identifying the current section
larg_ent	Width of the maintained pavement
nbvoi	Number of lanes

Appendix 1. Description of the 78 available variables (part 3 and end)

[•••]

G. 14. Pavement surface characteristics	
valcft	Coefficient of transversal friction, grip measurement (microtexture)
valhs	Gravel stone emergence, grip measurement (macrotexture)
valuni	Smoothness, measurement of pavement planeness
arrach	Ravelling, degradations of surface scaling type.
glac_res	Glazing / flushing, degradations of striping type.
deffaib	Slight deformations (ruts, slumps... of depths < 3 cm)
defgrav	Slight deformations (ruts, slumps... of depths > 3 cm)
rep_loc	Localised repairs Réfections ponctuelles of pavement (< ½ travelling lane)
rep_imp	Localised repairs Réfections ponctuelles of pavement (> ½ travelling lane)
ind_sur	Index of surface notation, aggregated indicator of previous elementary states, included between 0 and 20

[•••]

[•••]

G. 15. Structural pavement characteristics	
flfaib	Slight longitudinal cracks
flgrav	Severe longitudinal cracks
ftfaib	Slight transversal cracks
ftgrav	Severe transversal cracks
ind_pat	Indice patrimoine aggregated indicator of the previous elementary states, included between 0 and 20

G. 16. Nature of the pavement : types of material used for roadway surfacing.	
BBC	Nailed bitumen concrete
BBDR	Draining bitumen concrete
BBM	Thin bitumen concrete
BBTM	Very thin bitumen concrete
BBSG	Partially grained bitumen concrete
ES	Surface dressing
T	Special technique
ENROB	Coated material

G. 17. Nature of bonding used for surface layer : material used to bind the gravel forming the pavement coating.	
GB	bitumen-bound graded aggregate
GH	Cement-bond aggregate
NT	Untreated

END of chart

Appendix 2. Accident probability and conditional accident severity results

	variable name	model prob:1	model prob:2	model grav:1	model grav:2
		Accident probability*(a)*		**Probability of severity *(g):*** *Slight injuries = (B .L.)* *Severe injuries = (B.G.)* *Fatal = (M.)*	
I Probability points; Student's t *II **Value of λ ; Student's t***		*π (a)* ***(t =0)*** *λ [t =0;t=1]*	*π (a)* ***(t =0)*** *λ [t =0;t=1]*	*π (g)* ***(t =0)*** *λ [t =0;t=1]*	*π (g)* ***(t =0)*** *λ [t =0;t=1]*
G 1. Speed					
Expected standard deviation of speed	*et*		.001 (.03)		.029 .023 (.01) -.052 (-1.31)
Expected mean speed *LAMBDA*	*vml* *LAM 4*		-.202 (-1.63) 4.233[1.73][1.32]		-.247 .138 (1.37) .108 (1.54) 1.802 [.29][.13]
G 3. Traffic volume					
Average annual daily traffic all vehicles *LAMBDA*	*MJA* *LAM 1*	.114 (4.00) .535 [2.77][-2.40]	.101 (3.11) .525 [2.56][-2.31]		
Heavy goods vehicles percentage	*PourcPL*	-.107 (-3.81)	-.110 (-3.66)	-.047 -.030 (-.06) .077 (2.81)	-.025 -.048 (-.33) .073 (2.45)
G 5. Approaching zone in length (reference to descent)					
Non-stop climb	*montée*	-.035 (-1.25)	-.033 (-1.11)	.032 .013 (-.15) -.045 (-1.84)	.041 .008 (-.29) -.049 (-2.06)
Altered profile	*profalt*	-.006 (-.24)	-.003 (-.13)	.058 -.005 (-.71) -.054 (-2.63)	.063 -.010 (-.84) -.053 (-2.66)
Unknown profile	*lginc*	-.005 (-.10)	-.002 (-.03)	.026 -.027 (-.35) .001 (-.08)	.035 -.032 (-.44) -.003 (-.21)
G 6. Approaching horizontal zone (reference to practically straight alignment)					
Practically straight followed by a turn	*qdqdv*	-.015 (-.58)	-.019 (-.67)	-.124 .130 (3.03) -.006 (.56)	-.129 .141 (3.02) -.011 (.39)
Practically straight followed by a two-turn sequence	*qdvv*	-.021 (-.48)	-.028 (-.59)	-1.04 .075 (1.24) .030 (1.06)	-.112 .089 (.136) .023 (.92)
Two-turn sequence followed by a practically straight layout	*virvirqd*	-.002 (-.05)	.004 (.08)	.136 -.006 (-.90) -.131 (-2.39)	.137 -.016 (-.99) -.122 (-2.34)
Turn, practically straight alignment and turn sequence	*vqdv*	0.006 (.15)	.008 (.19)	.033 -.083 (-.94) .050 (1.16)	.049 -.082 (-.97) .033 (.61)
Three-turn sequence	*virvirvir*	-0.69 (-1.62)	-.082 (-1.77)	-.127 .082 (1.48) .045 (1.53)	-.152 .113 (1.75) .040 (1.39)

Appendix 2. Accident probability and conditional accident severity results (p.2)

G 7. In length element (reference to climb)					
Descent	*drampe*	-.007 (-.28)	-.006 (-.22)	.056 -.015 (-.72) -.042 (-1.82)	.058 -.013 (-.71) -.045 (-1.95)
Positive curvature parabola	*parbos*	.011 (.29)	.006 (.15)	-.080 .111 (1.52) -.031 (-.38)	-.083 .114 (1.56) -.031 (-.38)
Negative curvature parabola	*parcre*	.037 (1.18)	.034 (1.03)	-.084 .038 (1.11) .046 (1.93)	-.101 .047 (1.32) .054 (2.23)
Positive curvature connection	*racbos*	-.129 (-2.01)	-.144 (-2.12)	-.172 .090 (1.09) .082 (1.82)	-.181 .096 (1.15) .085 (1.94)
Negative curvature connection	*raccre*	-.016 (-.37)	-.021 (-.47)	-.163 .128 (1.93) .035 (1.44)	-.187 .140 (2.12) .047 (1.76)
G 8. Horizontal element (reference to the right)					
Circle to the right	*cerdro*	.004 (.13)	.004 (.12)		
Circle to the left	*cergau*	.029 (1.04)	.031 (1.05)		
Connection to the right	*racdro*	-.163 (-3.67)	-.168 (-3.60)		
Connection to the left	*racgau*	-.063 (-1.72)	-.068 (-1.75)		
G 9. Difficulty associated with the in length element					
Degree of slope or of ramp	*pen_ramp*			.004 .001 (.19) -.005 (2.01)	.005 .001 (.28) -.006 (2.29)
G 10. Difficulty associated with the horizontal element					
Heading angle *LAMBDA*	*anglecap* *LAM 2*	.044 (2.51) 2.439 [1.42][.84]	.051 (2.62) 2.24 [1.39][.77]	.043 -.082 (-2.08) .039 (1.44)	.054 -.086 (-2.24) .032 (1.00)
G 11. Hierarchy (reference to national road*)*					
Land development motorway links	*GLAT*	-.074 (-1.34)	.013 (-.16)	.072 -.159 (-1.28) .086 (1.51)	.144 -.222 (-1.69) .079 (.84)
Links ensuring completion of the motorway network	*LACRA*	-.148 (-2.88)	-.093 (-1.97)	-.051 -.061 (-.10) .112 (1.62)	.001 -.116 (-.47) . 115 (1.38)
Trunk roads	*RNL*	.002 (.09)	.007 (.24)	.054 -.048 (-1.18) -.006 (-.55)	.051 -.045 (-1.11) -.006 (-.54)
Urban express roads	*VRU*	.110 (.98)	.161 (1.28)	.620 -.630 (-3.97) .010 (-1.07)	.583 -.641 (-3.78) .059 (-.45)
G 12. Profile class (reference to two lanes)					
Dual two-lane	*deux_2v*			-.048 .197 (1.00) -.149 (-1.78)	-.067 .211 (1.12) -.144 (-1.69)
Four lanes	*quatrev*			.015 -.098 (-.75) .083 (1.84)	.005 -.094 (-.66) .088 (2.03)
Three lanes	*troisv*			-.017 .026 (.42) -.010 (-.21)	-.031 .040 (.67) -.009 (-.10)

Appendix 2. Accident probability and conditional accident severity results (p.3)

G 13. Road variable					
Length of the current section *LAMBDA*	*long* *LAM 1*	.072 (6.22) .535 [2.77][-2.40]	.072 (5.86) .525 [2.56][2.31]	-.039 .008 (1.10) .031 (3.27)	-.045 .011 (1.32) .035 (3.59)
G 14 et G 15. Structural and surface pavement characteristics					
Smoothness *LAMBDA* *Associated binary*	*valuni* *LAM 1* *bvaluni*	-.070 (-1.42) .535 [2.77][-2.40] .074 (1.23)	-.035 (-.60) .525 [2.56][-2.31] .073 (1.12)		
Variable A	***A***	-.054 (-2.11)	-.057 (-2.08)		
Variable B *LAM1* *Associated binary*	***B*** ***dB***			.005 -.011 (-1.12) .007 (1.08) 1.691 [2.24][.97] -.008 .070 (1.08) -.062 (-2.29)	.006 -.016 (-1.05) .010 (1.19) 1.480 [2.51][.81] -.010 .072 (1.09) -.062 (-2.27)
Variable C	***C***			.086 -.092 (-2.42) .006 (-.39)	.087 -.097 (-2.48) .009 (-.27)
Variable D *LAMBDA* *Associated binary*	***D*** ***LAM1*** ***dD***			.022 -.000 (-1.28) -.021 (-1.94) 1.691 [2.24][.97] -.049 .051 (.95) -.002 (.17)	.041 -.005 (-1.51) -.036 (-2.14) 1.480 [2.51][.81] -.086 .074 (1.37) .012 (.68)
Variable E *LAMBDA* *Associated binary*	***E*** ***LAM2 /*** ***LAM1*** ***dE***	-.000 (-1.00) 2.439 [1.42] [.84] .144 (1.92)	-.000 (-.86) 2.240 [1.39] [.77] .138 (1.72)	.012 -.005 (-1.10) -.007 (-0.71) 1.691 [2.24][.97] -.252 .087 (1.20) .165 (2.58)	.030 -.016 (-1.24) -.014 (-.77) 1.480 [2.51][.81] -.283 .117 (1.28) .166 (2.59)
Variable F *LAMBDA* *Associated binary*	***F*** ***LAM1*** ***dF***			-.028 .019 (3.54) .009 (3.82) 1.691 [2.24][.97] .148 -.077 (-2.12) -.071 (-2.34)	-.051 .034 (3.68) .017 (3.97) 1.480 [2.51][.81] .154 -.085 (-2.21) -.069 (-2.33)
Variable G *LAMBDA* *Associated binary*	***G*** ***LAM3 /*** ***LAM1*** ***dG***	-.763 (-1.68) -.341 [-.31] [-1.23] .124 (2.55)	-.909 (-1.47) -.516 [-.35] [-1.03] .129 (2.42)	.003 -.003 (-1.05) -.000 (-.47) 1.691 [2.24][.97] .031 .061 (.42) -.092 (-2.07)	.009 -.008 (-1.28) -.001 (-.63) 1.480 [2.51][.81] .012 .072 (.67) -.085 (-1.84)

Annexe 2. Accident probability and conditional accident severity results (p.4)

Variable H	***H***	.013 (.86)	.010 (.78)	-.030	-.037
				.036 (1.98)	.043 (2.02)
				-.006 (-.09)	-.006 (.02)
LAMBDA	***LAM3 /***	-.341 [-.31] [-1.23]	-.517 [-.35] [-1.03]	1.691 [2.24][.97]	1.480 [2.51][.81]
Associated	***LAM1***	-.006 (-.13)	-.012 (-.21)	-.014	-.012
binary	***dH***			-.017 (-.10)	-.020 (-.15)
				.031 (-.15)	.032 (1.40)
Variable I	***I***			-.010	-.018
				.008 (1.38)	.013 (1.59)
				.002 (1.20)	.005 (1.62)
LAMBDA	***LAM1***			1.691 [2.24][.97]	1.480 [2.51][.81]
Associated	***dI***			.026	.027
binary				.014 (-.10)	.008 (-.20)
				-.041 (-2.13)	-.036 (-1.90)

III. General Statistics				
Log-likelihood. Final	-725.669	-724.371	-1064.15	-1061.46
Log-likelihood with constant only	-847.365	-847.365	-1195.99	-1195.99
Rho squared	.144	.145	.110	.112
Rho carré bar				
Akaike	.099	.097	.043	.041
Horowitz	.121	.121	.076	.077
Hensher and Johnson	.131	.131	.080	.080
Percentage of good predictions	55.057	55.293	57.143	57.796
Number of alternatives	2	2	3	3
Number of observations	2541	2541	1225	1225
Number of estimated parameters				
Bêtas - variables	30	32	66	68
- constants	1	1	2	2
- associated binary	4	4	14	14
- Box-Cox transformations	3	4	1	1

Note : the original chart will be available when SETRA will authorise its diffusion.

Appendix 3. Associated binaries with variables containing zero observations

We have previously stated that we had used Box-Cox transformations (BC) on variables containing a portion of zero observations, in particular structure and surface variables of groups 14 and 15 concealed in Appendix 2. We want to specify here that these binary variables associated with non strictly positive variables play an essential role explaining that the TRIO software generates them automatically. Yet, since they modify the sense of the model, we must establish it with a simple well known linear case.

Binary variable of interaction in linear regression. Let's consider a typical linear regression model where the dependent variable ***y*** depends of a constant, of a ***Q*** variable that is not strictly positive, for example snowfall, and of a ***D*** associated binary variable, by this we mean adding the ***t*** index of the observations :

$$y_t = \beta_0 + \beta Q_t + \beta_d D_t + u_t \tag{25}$$

where $D_t = 1$ if $Q_t \neq 0$ and $D_t = 0$ otherwise. In this model, the D_t variable can be defined simply as a variable of interaction adding the effect of a state to the one of the quantity represented by Q_t : we would measure in our example the effect of the presence of snow in addition to the one of the snow quantity. Of course, if we define the state variable by its complement, $D_t = 0$ if $Q_t \neq 0$ and $D_t = 1$ otherwise, we only have to inverse the sign of β_d. Adding such a state variable eliminates a degree of freedom, modifies the formulated model without it, and might brings nothing to the explanation.

Binary variable and invariance in presence of BC transformations. If the model happens to be :

$$y_t = \beta_0 + \beta Q_t^{(\lambda)} + \beta_d D_t + u_t, \tag{26}$$

where the BC transformation defined in (13)-(14) is applied only to the positive values of Q_t, the associated binary variable will play another role : to preserve the invariance of the λ estimation of the measurement units used for Q_t. The analytic proof of this affirmation is hard to demonstrate, as opposed to the numeric proof, accessible if we have a regression software. The latter has the following structure. Let's fix λ to an arbitrary value, 2 for example, and apply the estimator of the ordinary least squares by conserving Q_t in original units. We obtain the vector $\hat{\beta}_0, \hat{\beta}, \hat{\beta}_d$. Let's estimated again with the scale factor s : $\tilde{Q}_t = sQ_t$. We obtain $\tilde{\beta}_0, \tilde{\beta}, \tilde{\beta}_d$ and we observe that $\hat{\beta}_0 = \tilde{\beta}_0$, that $\hat{\beta} = s^{\lambda}\tilde{\beta}$, and that $\hat{\beta}_d = \tilde{\beta}_d + s^{(\lambda)}\tilde{\beta}$. The adjustment needed by the absence of transformation of zero observations of Q_t is done by the coefficient of the associated binary variable. The model will be invariant because the coefficient of the constant cannot be adjusted while the coefficient of Q_t is adjusted as usual in an inversely proportional way to the factor s^{λ}. The associated binary variable has simultaneously two meanings here : a variable of interaction and a compensatory variable.

Invariance during the transformation of a strictly positive variable. We must not confuse this observed invariance with the one described by Schlesselman (1971) mentioned above in the case of variables of strictly positive values subjected to BC transformations. In that case, in presence of a constant of regression, its coefficient adjusts to a change of scale ***s*** of the variable Q transformed according to the following simple mechanism :

$$\beta_0 + \beta\,(Q^{\lambda} - 1)/\lambda) \tag{27}$$

$$\beta_0 + \beta s^{\lambda} \left(Q/s\right)^{\lambda} \Big/ \lambda - \beta/\lambda \tag{28}$$

$$\beta_0 + \widetilde{\beta}\, \widetilde{Q}^{\lambda}/\lambda - \beta/\lambda + \widetilde{\beta}/\lambda - \widetilde{\beta}/\lambda \tag{29}$$

$$\beta_0 + \widetilde{\beta}\, (\widetilde{Q}^{\lambda} - 1)/\lambda - \beta/\lambda + \widetilde{\beta}/\lambda \tag{30}$$

$$\widetilde{\beta}_0 + \widetilde{\beta}\, (\widetilde{Q}^{\lambda} - 1)/\lambda \tag{31}$$

where $\widetilde{\beta} = s^{\lambda}\beta$, such as in the previous case, where it is easy to analytically derive the new coefficient of the constant $\widetilde{\beta}_0 = \beta_0 - \beta/\lambda + \widetilde{\beta}/\lambda$.