Pricing of Automobile Insurance Under Asymmetric Information: a Study on Panel Data

by Maki Dahchour and Georges Dionne

Working Paper 01-06
May 2002

ISSN : 1206-3304

Earlier versions of this paper benefited from comments made by Jean Pinquet, Denis Bolduc, and Sylvain Lassarre; by participants at the "Sécurité routière et assurance automobile" symposium—FFSA (Paris, 18 December 2000) as well as those attending the 41st Congress of the Société Canadienne de Science Économique (Quebec, 14 and 15 May 2001). Financial support by INRETS, CRSH, and FFSA is acknowledged.

Pricing of Automobile Insurance Under Asymmetric Information: a Study on Panel Data

Maki Dahchour
Georges Dionne

Maki Dahchour is Ph D student at THEMA, Université Paris X, DERA-INRETS and CRT, University of Montreal.

Georges Dionne is Chairholder, Risk Management Chair, and professor, Finance Department, École des HEC Montreal.

Pricing of Automobile Insurance Under Asymmetric Information: a Study on Panel Data

Abstract

This article proposes to highlight the informational content of the French *bonus-malus* scheme used in *a posteriori* pricing and to verify whether the automobile insurance pricing scheme is efficient in eliminating all residual asymmetric information in the risk categories constructed by insurers. The article expands the asymmetric information test developed by Dionne, Gouriéroux, and Vanasse (2001) to panel data. The data are from the *Parc Automobile Sofres* in France. An incomplete panel composed of 11,506 individuals was constructed over a three-year period (1995–1997). We show that the variables used by insurers in pricing automobile insurance in France efficiently account for asymmetric information. Moreover, the *bonus-malus* variable turns out to be significant in explaining both the individual distribution of accidents and the type of insurance coverage chosen. However, its absence does not affect the conclusion about the presence of residual asymmetric information.

Keywords: Asymmetric information, automobile insurance, *bonus-malus*, risk classification, panel data, state dependence, unobserved heterogeneity, random-effects probit, negative binomial distribution, road safety.

JEL Numbers: D80, G22, C23, L51.

Résumé

Cet article propose d'étudier la valeur informationnelle du *bonus-malus* français utilisé dans la tarification a posteriori de l'assurance automobile et de vérifier comment la tarification est efficace pour éliminer toute forme d'asymétrie d'information résiduelle dans les classes de risque créées par les assureurs. L'article étend le test d'asymétrie d'information développé par Dionne, Gouriéroux et Vanasse (2001) à des données de panel. Les données proviennent du "Parc Automobile" Sofres en France. Un panel incomplet de 11 506 individus a été construit sur la période 1995-1997. Nous montrons que les variables utilisées par les assureurs sont efficaces pour éliminer l'asymétrie d'information résiduelle. De plus, la variable *bonus-malus* est significative pour expliquer les taux d'accidents et les choix de couverture d'assurance. Par contre, elle n'est pas nécessaire pour conclure sur l'absence ou non d'asymétrie d'information résiduelle dans les portefeuilles des assureurs.

Mots clés: information asymétrique, assurance automobile, *bonus-malus*, classification des risques, données de panel, dépendance d'état, hétérogénéité inobservée, probit à effets aléatoires, distribution binomiale négative, sécurité routière.

Classification JEL : D80, G22, C23, L51.

Introduction

In France, automobile insurance pricing is based on two elements. The first is a so-called *a priori* pricing system which consists in constructing classes of homogenous risks based on the characteristics of policy-holders. The second is an *a posteriori* pricing mechanism called *bonus-malus*, where past accidents are used to fix the premium in the next period. The purpose of the latter is to modulate individual premiums in terms of past accidents and thus correct the imperfections inherent in *a priori* pricing. The *bonus-malus* is also used to offer incentives for safe driving. Application of the *bonus-malus* scheme was compulsory in France during the period of our analysis.

There is full commitment by the insurance industry on the application of the *bonus-malus* scheme. This commitment is enforced by a law. This implies that each insurer must apply the same formula for premiums for the same kinds of driving history according to ex-ante rules that cannot be renegotiated. The European Commission is against such regulation but has not considered the value of information related to the application of the *bonus-malus* scheme in France. It is well known that a *bonus-malus* scheme can be complement to ex-ante risk classification for taking into account of both moral hazard and adverse selection (Dionne, 2002). So it is important to study the informational content of this *bonus-malus* scheme.

We expand the Dionne, Gouriéroux, Vanasse (DGV, 2001) asymmetric information test to take into account of panel data modeling. This obviously entails adapting the DGV test to this type of data which reflects the dynamic nature of the decisions and behaviors of drivers. What we want to know in particular is whether the French *a posteriori* (or *bonus-malus*) pricing system can effectively take individual risks into account. Special attention will thus be paid to the informational content of the *bonus-malus*, which is supposed to contain the full history of each driver’s accident profile. We want to see whether the *bonus-malus* brings any additional information to bear in explaining drivers’ choice of insurance coverage as it relates to the record of accidents observed during the period over which the panel is constructed and to the estimates of accidents predicted for each of the periods.

Our analysis is based on an incomplete panel which changes composition as drivers enter and leave the panel during the 1995–1997 period. Our database contains 16,399 observations, corresponding to 11,506 vehicles-drivers.

The article is organized in six parts. In the first part, we present the methodology of the DGV test (2001). In the second, we give the characteristics of the French automobile insurance market. The third to fifth parts present, respectively, our objectives, the database, and the asymmetric information test for longitudinal or panel data. The sixth part analyses the results of our various regressions. Finally, our conclusion reviews the main findings and their implications.

1. Methodology

The residual asymmetric information test is based on the notion of conditional independence. The endogenous variable is denoted as Y , the exogenous k variables as X and the p variables for the decisions of individuals (here drivers) as Z. In our case, X may contain the driver's age and sex and the characteristics of the vehicle. Y may be the number of accidents or traffic violations the policyholder has been involved in over the year and Z may be the deductible or insurance coverage (third-party or all-risk insurance) chosen when the policy was taken out.

There is, by definition, conditional independence when the probability distribution of Y, as conditionally determined by both the vector X of explanatory variables and the vector Z of decisional variables, coincides with the probability of Y distribution as determined solely by the vector X of explanatory variables and vice versa. This would mean that, given X, knowledge of the Z variables will add no information to the knowledge of Y. Formally, this translates into the following relations:

$$f(Y|X,Z) = f(Y|X)$$

which is equivalent to

$$f(Z|X,Y) = f(Z|X)$$

where $f\left(\cdot\,|\,\cdot\right)$ is a conditional density function.

Dionne, Gouriéroux and Vanasse (1998, 2001) proposed a methodology based on the above relationships to test for asymmetric information in the portfolio of a private insurer while Dionne and Gagné (2002) proposed a method, based on the same relationships, to separate moral hazard from adverse selection.

To accomplish this task, Dionne, Gouriéroux and Vanasse estimated two equations: the first uses a negative binomial model to estimate the expected number of accidents, whereas the second uses a probit model to estimate the choice of deductible. The second equation introduces the concept of conditional dependence to show that, frequently, the presence of asymmetric information is not rejected when the non-linearity of insurance pricing is not taken into account. For the authors, the "right" test is the one which would use, as explanatory variable for the deductible choice, simultaneously the expected number of accidents (determined based on modeling the conditional distribution of accidents, thus on explanatory variables) and the actual number of accidents. This produces results revealing that the actual number of accidents is of no significance, thus allowing the authors to reject the presence of asymmetric information. In point of fact, the actual number of accidents is not significant in the portfolio of the studied insurer, meaning that, when properly classified, the policy-holder will not be better informed about his type of risk than the insurer. The fact that the "estimated expected number of accidents" variable is significant indicates that the econometric modeling still contains non-linearities. In a last step, Dionne, Gouriéroux, and Vanasse show that it is possible to render this variable non-significant by adding interactions between the risk-classification variables (see Chiappori and Salanié (2000) for an equivalent test and Finkelstein and Poterba (2000) for a test on the presence of adverse selection).

In our case, we had no variable for the choice of deductible in our database for 1995, 1996, and 1997. However, choice of type of insurance coverage can be a substitute for this variable (Chiappori and Salanié, 1997, 2002).

2. The French automobile insurance market

Several statements can be made based on observation of the French automobile insurance market:[1]

- There is no clause in insurance contracts obliging policy-holders to stay with the same insurance company once the contractual period is over. So, only semi-commitment is possible in this market, as for many insurance markets. (On the notion of semi-commitment, see Dionne and Doherty, 1994, and Hendel and Lizzeri, 2000).
- If the accident is the driver's fault, this is public information insofar as rival companies will be informed of the fact. This information is given in the report that the subscriber must provide when purchasing a new insurance contract. Moreover, there is full commitment by the industry on the application of the *bonus-malus* scheme.
- Information (type of coverage and premium paid) about any previous contracts purchased is usually public.
- Exclusivity of contract purchases (only one insurer).
- Low-risk clients subsidize high-risk clients: the most often cited perverse effect generated by the French *bonus-malus* system is the artificial inflation of insurance premiums for some risk classes. This inflation is more specifically a concern for young drivers.

These observations tend towards of any commitment on the policyholder's part and imply public information between insurers on contract parameters, two ingredients which characterize the form of a long-term contracting (Dionne, Doherty and Fombaron, 2000).

But the existence of the Reduction-Increase clause (as the *bonus-malus* is officially designated in France)[2], the only one sanctioned by law, advocates for full commitment on the part of the

[1] See, for example, Richaudeau (1998) and Fombaron (2002) for a detailed description of the automobile insurance market in France and Pinquet (1999) for an analysis of the *bonus-malus*.
[2] In France, according to the regulations in force (Article A 121-1 - Automobile insurance - Reduction-Increase clause of the French Insurance Code and its appendix), the premium charged the policy holder is necessarily determined by multiplying the amount of the actuarial premium by a so-called "reduction-increase coefficient." The base coefficient is 1. After each year of insurance without accident, the coefficient used is that of the preceding

industry. It is as if there were a dynamic agreement between the policyholder and the insurance industry. The insurers make a commitment regarding the future evolution of premiums since the *a posteriori* pricing formula is known at the inception of the contractual relationship and is legislatively regulated to remain stable (cf. G Rosenwald, 2000; Dionne, 2002). This commitment regarding the *bonus-malus* has two advantages. On the one hand, it puts some teeth in the threat of sanctions against drivers who have accidents. On the other hand, the “carrot” of reduced premiums in the next period promised to accident-free drivers is an incentive to the practice of preventive driving.[3]

Theoretical studies have shown that, if there is no form of commitment on the part of the insurer, the *bonus-malus* will offer no incentive whatsoever (Dionne et al., 2000). Similarly, if a particular insurer benefits from an informational edge in the form of private data on his client’s accidents, this will eliminate any benefits from the *bonus-malus,* because insureds will choose another insurer when the malus will increase (Kunreuther and Pauly, 1985; Fombaron, 1997).

Consequently, we conclude that there is a commitment, in France, regarding the *bonus-malus* on the industry’s side and that there is no form of commitment on the other pricing elements, particularly when a policyholder leaves his insurance company for another (strong competition).

3. Objectives

The purpose of this paper is to answer the following questions:

(i) Will the conclusions derived from tests conducted using *cross-sectional* data on the optimality of automobile insurance pricing in France (Richaudeau, 1999; Chiappori and Salanié, 2000) retain their validity with modeling on longitudinal data?

---

contract period minus 5%. Each accident occurring in the insurance year increases the coefficient by 25%. The clause also stipulates that there is no increase for the first accident occurring after a period of at least three years during which the reduction-increase coefficient was equal to 0.50.

[3] See P. Picard (2000) for an analysis of the deregulation of insurance markets in Europe.

(ii) What explanatory power do variables such as *bonus-malus* and mileage have in the dynamic model for insurance coverage?

(iii) Is the *bonus-malus* variable necessary for the test of residual asymmetric information in risk classes?

These three questions lead us to the following econometric problems:

- What is the best specification for the insurance coverage model when dealing with panel data: fixed or random effects?
- What is the best combination of explanatory variables for explaining individual risk if longitudinal data are used to account for the dynamics?
- What form must the asymmetric information test take when using panel data?

The answer to the first problem led to a random-effects specification of the insurance coverage based on panel data integrating *bonus-malus*, exposure to risk and risk classification variables as the main explanatory variables (see Pinquet, 2000 and Dahchour and Lassarre, 2001 on panel analysis of accidents distribution). For the next three, we propose the following econometric hypotheses:

- As for the *cross-sectional* data, adequate risk classification should suffice to account for individual risks in a dynamic modeling of the choice of insurance coverage.
- *Bonus-malus* should affect the choice of insurance coverage. The *bonus-malus* can be expected to correlate negatively with the purchase of all-risk insurance coverage.
- The *bonus-malus* coefficient should also be a predictor of a driver's risk of having an accident.
- However, it is not clear that the *bonus-malus* variable will be a necessary ingredient to test for the presence of asymmetrical information.

4. Panel data

Individual panel data or longitudinal data have several advantages over aggregate data:[4] elimination of any bias linked to aggregation and hence more clearly defined estimators; more accurate measurement of certain variables, explicit measure of individual heterogeneity allowing the researcher to go beyond the notion of representative agent towards the search for better econometric modeling capable of encompassing it. In the specific case of the automobile insurance data with which we are here concerned, the longitudinal approach can also highlight changes in drivers' behavior and characteristics (choice of insurance coverage, for example) from one period to the next, following the occurrence of a traffic accident. These changes may be qualified as endogenous since they are linked to the driver himself. Analysis based on this type of data thus makes it possible to account for the advantages linked to these changes to capture the dynamic effect of the accident event and thus compensate for the flaw inherent in considering these data as *cross-sectional* data. However, changes in a driver's behavior may be motivated by reasons other than involvement in an accident. Other short- or long-term influences on the behavior of drivers may include: change in legislation concerning the traffic code; modification of automobile-insurance-pricing policies either by insurance companies or by regulatory authorities (for example, modification of the standard clause related to *bonus-malus*); launch of an accident prevention campaign; or modifications in the road network. These would thus be exogenous influences since they are linked to external events. So it is very difficult to distinguish that share in the modification of a driver's risk status which can be attributed to a change in behavior following an accident from that linked to other external influences. If we look at the period covered by our panel (1995–1997), there was, to our knowledge, no event of the types cited above, at least none linked to the first three types mentioned.

During construction of the panel, we were faced with three problems which caused the loss of a great deal of information and eliminated observations at various steps:

(i) Entries to and exits from the panel: about a third of the Sofres panel turns over each year. These entries and exits are attributable to the following constraints:

- Sofres' concern to fairly represent the French population;

[4] See, for example, Diggle, Liang, and Zeger (1994) and Mátyàs and Sevestre (1999).

- elimination of households having refused to answer previous surveys;
- change of residence, death…

(ii) With each survey, observations were eliminated because questions crucial to our analysis went unanswered: for example, questions concerning the *bonus-malus* and the type of insurance coverage chosen, to mention only the most important.

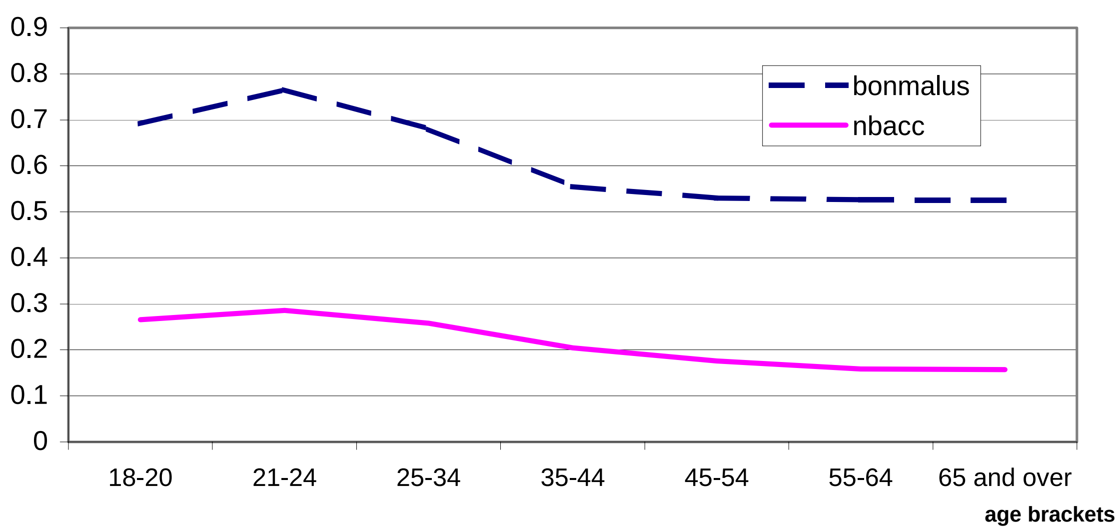
(iii) When cross-matching to construct the panel, the main difficulty was a lack of any key or tracer allowing us to "follow" a vehicle from one period to the next. The only tracer available and informatively linked to all the observations had to do not with the vehicle but with the household; but a household can own several cars. This is the reason why we started looking for a better combination of criteria (or variables) which would allow us to recuperate the largest number of vehicles possible from one survey to the next. The ideal would have been to combine the household and the vehicle model tracers to cross-check but, unfortunately, the latter is missing for a lot of vehicles. Thus, after several attempts using the household tracer and various other criteria (year of vehicle, the first four digits of the license number, year driver's license, category of vehicle, buying power, price category), we opted for the combination identifying the household and the first four digits of the license number. This is the pair of criteria which allowed us to construct the most reliable panel in terms of quality the cross-matching.

Our final sample of data is thus a three-period panel (1995–1997). The panel is incomplete; in the sense that the same individual-vehicle is not necessarily present in each period. It is composed of 16,399 observations corresponding to 11,506 individuals-vehicles. The resulting *balanced* panel contains 1,106 individuals-vehicles: about 10% of the incomplete panel.

The information available in the database is composed of three elements. The first concerns information on driver characteristics (sex, age, number of accidents…). The second covers the vehicles (year, group, …). The third and most important element for the problem we are studying relates to insurance contracts. Unfortunately, this informational element is very limited, containing nothing about insurance premiums or deductibles, let alone anything about switching insurers from year to year. It provides, however, the *bonus-malus* coefficient and the type of insurance coverage.

If we look at the evolution of the average *bonus-malus* and the average frequency of accidents in terms of the age of the principal driver, we notice from Figure 1 that both variables show the same trend, though the accident rate seems less volatile. This is however a sign that the *bonus-malus* is tightly indexed to past accidents and that it probably contains the same information as that provided by accidents. It should be noted that only accidents over the three years covered by the panel (1995, 1996, and 1997) are of concern here, whereas the *bonus-malus* coefficient is supposed to contain the whole history of accidents.

Figure 1: Evolution of average *bonus-malus* and of accident rate, by age



A second graph represents the evolution of both the average accident rate and the average *bonus-malus* according to age brackets and choice of insurance coverage (all-risk or third-party insurance) (Figure 2). We note that even if drivers having chosen all-risk insurance do, on average, have more accidents than drivers with third-party insurance, the former still show a lower average *bonus-malus* coefficient. The same pattern is noted when observing the evolution of these two variables over the period studied (1995-1996-1997) (Figure 3). We were rather expecting the two curves to have levels of the same order.

Figure 2: Evolution of observed accident rate and *bonus-malus* by age bracket

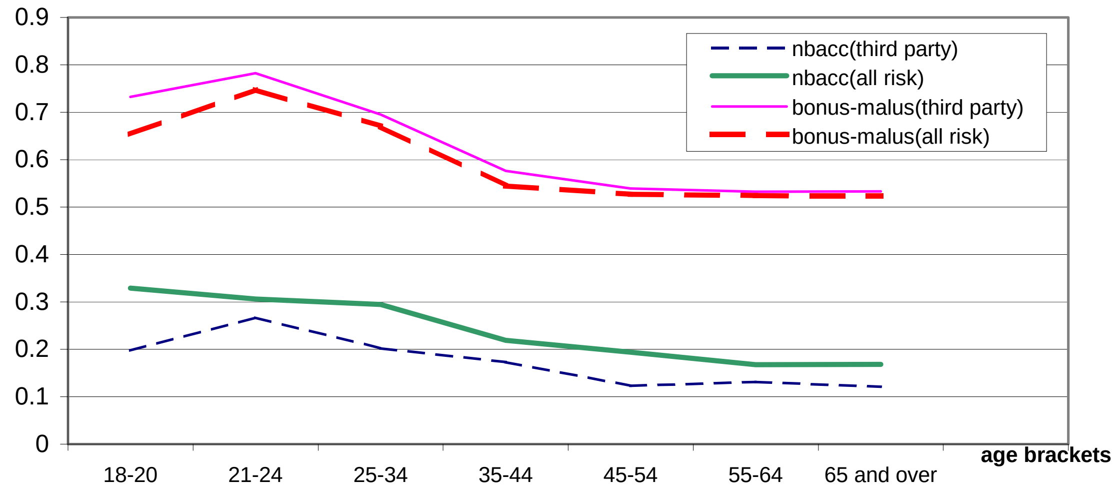
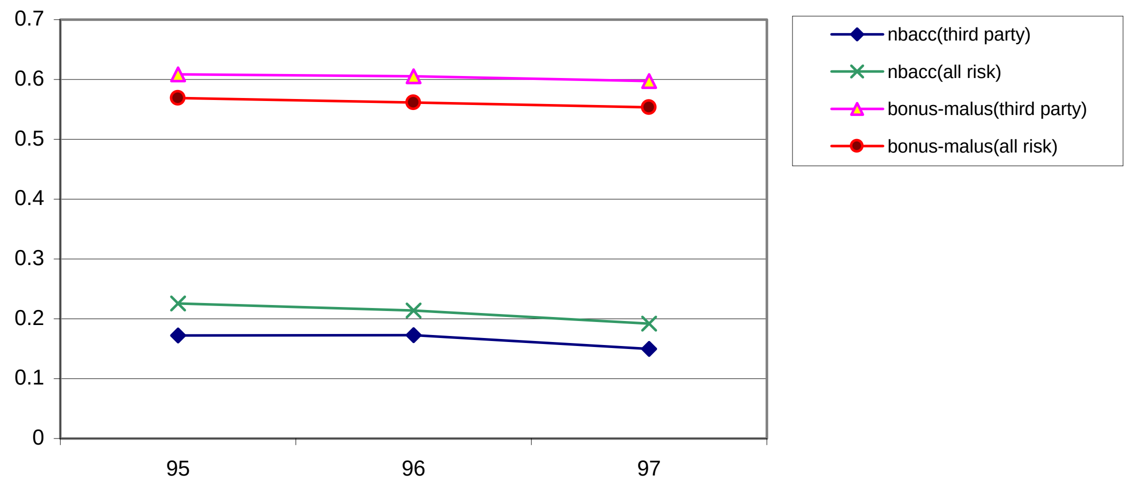


Figure 3: Evolution of accident rate and bonus-malus over the period, according to type of insurance coverage



5. Test equations

As already pointed out in the introduction, our objective is to expand the DGV test for residual asymmetric information (or optimality of insurance pricing) in order to treat individual, temporal data (or panel data) while also taking the *bonus-malus* into account. In other words, we want to find out whether or not insurers can carry out effective risk classification relying on policy-holders' characteristics and on experience-based pricing.

The test proposed here is based on a two-step estimation. The first concerns predictions of the expected number of accidents during each period, whereas the second deals with the choice between all-risk and third-party insurance coverage, to verify whether this choice is explained by private information. In the second step, we incorporate into the list of explanatory variables the "expected number of accidents" of each year obtained with the first estimation. This type of estimation is known to produce convergent estimators in the second step (Murphy and Topel, 1985; Greene, 2000).

There are thus two equations to estimate: The first is designed to estimate, in each period, the individual risk as conditioned by the explanatory variables, thus producing the accident forecast. In our case, the appropriate tool for this is the negative binomial model. The second equation integrates the information derived from the first equation into its regression components in order to estimate the choice of insurance coverage (all-risk vs third-party insurance). The two equations will allow us to test whether or not the *bonus-malus* is effective through its informational content. We want to see if the *bonus-malus* adds any information to the knowledge of risks over and above that derived from the estimation of the second equation. Finally, we want to test if the *bonus-malus* affects the conclusion on the presence of residual asymmetrical information.

Equation 1: Expected number of accidents: the negative binomial model

We assume that the number of accidents Y_i in which an individual i is involved in a given period will follow a negative binomial law:

$$P(Y_i / X_i, \alpha, \beta) = \frac{\Gamma(\alpha + Y_i)}{\Gamma(\alpha)\Gamma(Y_i + 1)} \frac{\left(\alpha^{-1} \exp(X_i \beta)\right)^{Y_i}}{\left(1 + \alpha^{-1} \exp(X_i \beta)\right)^{Y_i + \alpha}}$$

where $\Gamma(.)$ is the gamma function such that $\Gamma(z) = t^{z-1} e^{-t}$ for $z > 0$, X is the vector of explanatory risk-of-accident variables (age, sex,…), β is the vector (of appropriate size) of the parameters to be estimated, α is the model's overdispersion parameter.

The estimation of β will allow us to obtain the expected number of accidents for 1995, 1996, and 1997. We note *E(nbacc95), E(nbacc96), E(nbacc97),* respectively the accidents predicted by the three models estimated for 1995, 1996, and 1997. When the longitudinal data are considered, they form an explanatory variable over time, noted *E(NBACC),* which will be introduced into the second equation. (See Table 7 in the appendix for the results of the negative binomial model over 1995-1997). Note that the *bonus-malus* variable has a positive coefficient which means that bad drivers (high *bonus-malus*) have more accidents.

Equation 2: Choice of insurance coverage: all-risk vs third-party insurance: the probit panel model

- **Specification**

The "choice of insurance coverage" variable is binary: it is valued 1 if the policy holder chooses all-risk insurance and 0 if he opts for third-party insurance. All-risk insurance represents more coverage so it should be chosen by the high-risk individuals according to the Rothschild-Stiglitz (1976) model. Letting $\hat{Z}$ stand for the latent variable related to the utility associated with the choice of all-risk insurance over third-party insurance, *Z* for the binary variable of the choice of type of insurance coverage, and *X* the variables explaining this choice, we may write:

$$\hat{Z}_{it} = X_{it}b + \varepsilon_{it}, \; i = 1, \ldots, N \text{ and } t = 1, \ldots, T_i,$$

with

$$Z_{it} = 1 \quad \text{if} \quad \hat{Z}_{it} = X_{it}b + \varepsilon_{it} > 0$$

$$Z_{it} = 0 \quad \text{if} \quad \hat{Z}_{it} = X_{it}b + \varepsilon_{it} \leq 0$$

where *b* is the vector of the appropriately sized parameters to be estimated and ε_{it} the error term. If we assume that the latter follow a normal law, we end up with a traditional binomial probit model. Moreover, if we check the hypothesis that the errors are not correlated in time for the same individual (i.e., $Corr(\varepsilon_{it}, \varepsilon_{is}) = 0$), then the panel modeling techniques need not be used. However, whenever this hypothesis no longer applies (i.e., $Corr(\varepsilon_{it}, \varepsilon_{is}) \neq 0$) the panel methods must be taken into account. Two types of specification exist for panel data: random-effects specification and fixed-effects specification. The first assumes that specific individual effects are certain and vary with time, whereas the second assumes that these effects are "drawn" once for all and, for any given individual, will recur identically over time.

Just as when modeling accident rates on longitudinal data, the random-effects model is intuitively preferred for modeling choice of insurance coverage, particularly when the data set contain a large number of individuals. The random-effects probit model has undergone a number of developments (Heckman, 1981; Butler and Moffit, 1982; Guilkey and Murphy, 1993; Hsiao, 1996; …).

Thus, when there are individual specific effects denoted u_i (considered random) and if errors are denoted as v_{it}, the preceding model becomes:

$$\hat{Z}_{it} = X_{it}b + u_i + v_{it}, \text{ for } i = 1,\dots,N \text{ et } t = 1,\dots,T_i .$$

Let

$$\varepsilon_{it} = u_i + v_{it},$$

with

$$\begin{pmatrix} u_i \\ v_{it} \end{pmatrix} \sim N\left(0, \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & 1 \end{pmatrix}\right).$$

This assumes that errors terms ε_{it} are independent from one individual to the next. However, it shall be assumed that they are normal and correlated over time for the same individual owing to the influence of u_i, so that variance ε_{it} is written:

$$V(\varepsilon_{it}) = V(u_i + v_{it}) = \sigma_v^2 + \sigma_u^2 = 1 + \sigma_u^2$$

and

$$Corr(\varepsilon_{it}, \varepsilon_{is}) = corr(u_i + v_{it}; u_i + v_{is}) = \rho = \frac{\sigma_u^2}{1 + \sigma_u^2}, \ for\ t \neq s .$$

Interpretation of the ρ correlation coefficient is direct and consists in saying that it measures the part of the variance arising from the individual effect in the error's total variance. This hypothesis of *temporal equicorrelation* of the individual effects u_i, will prove to be crucial in the estimation of the model.

Thus, our random-effects probit model for choice of insurance coverage is written as:

$$\begin{cases} Z_{it} = 1 \quad if \quad \hat{Z}_i = X_{it}b + u_i + v_{it} > 0 \quad \text{(if all-risk insurance is chosen)} \\ Z_{it} = 0 \quad if \quad \hat{Z}_{it} = X_{it}b + u_i + v_{it} \le 0 \quad \text{(if third-party insurance is chosen)} \\ with \quad b = \dfrac{b^*}{\sigma_u} \text{ and} \\ Corr(\varepsilon_{it}, \varepsilon_{is}) = \tilde{n} = \dfrac{\sigma_u^2}{1+\sigma_u^2}, \text{ for } t \ne s \text{ (or } \sigma_u^2 = \dfrac{\tilde{n}}{1-\tilde{n}}) \end{cases}$$

- **Estimation of the model**

To avoid using two separate terms corresponding to $Z_{it} = 1$ and to $Z_{it} = 0$, let $r_{it} = 2Z_{it} - 1$.

We then have:

$$P(Z_{it}) = \int_{-\infty}^{r_{it}X_{it}b} \phi(\varepsilon_{it}) d\varepsilon_{it} = \Phi(r_{it}X_{it}b).$$

The probability log related to individual *i* is obtained in the following manner:

$$L_i(b) = \sum_{i=1}^{N} \log[P(Z_{i1}, \ldots Z_{iT})]$$

with

$$P(Z_{i1}, \ldots Z_{iT}) = \int_{-\infty}^{r_{i1}X_{i1}b} \ldots \int_{-\infty}^{r_{iT_i}X_{iT_i}b} \phi(\varepsilon_{it}, \ldots, \varepsilon_{iT_i}) d\varepsilon_{it} \ldots d\varepsilon_{iT_i}.$$

Integration with the combined density is not practical, perhaps even impossible. However, a specification of the model in terms of random effects allows us to simplify the above notation.

Indeed, it is possible to make u_i conditional in the expression of combined density $\phi(\varepsilon_{it},\ldots,\varepsilon_{iT_i},u_i)$ since

$$\phi(\varepsilon_{it},\ldots,\varepsilon_{iT_i},u_i)=\phi(\varepsilon_{it},\ldots,\varepsilon_{iT_i}|u_i)\phi(u_i).$$

Then,

$$\phi(\varepsilon_{it},\ldots,\varepsilon_{iT_i})=\int_{-\infty}^{+\infty}\phi(\varepsilon_{it},\ldots,\varepsilon_{iT_i}|u_i)\phi(u_i)du_i$$

and depending on u_i, the ε_{it} become independent so that we can write:

$$\phi(\varepsilon_{it},\ldots,\varepsilon_{iT_i})=\int_{-\infty}^{+\infty}\prod_{t=1}^{T_i}\phi(\varepsilon_{it}|u_i)\phi(u_i)du_i .$$

By introducing the change in variable $h_i=\frac{u_i}{\sqrt{2}}$ suggested by Greene (1997) we can show that:

$$L_i(b)=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{\pi}}e^{-h_i^2}\prod_{t=1}^{T_i}\Phi(X_{it}b+\theta h_i)\phi(u_i)dh_i$$

with

$$\grave{e}=\sqrt{\frac{2\tilde{n}}{1-\tilde{n}}}.$$

This expression belongs to the type $\int_{-\infty}^{+\infty}\exp(-z^2)g(z)$ which integrates numerically.

Butler and Moffit [1982] proposed an evaluation of this integral using the Gaussian quadrature method based on Hermitian integration (on this subject, see Greene, 2000, for example):

$$\int_{-\infty}^{+\infty}\exp(-z^2)g(z)\simeq\sum_{j=1}^{J}w_j g(z_j).$$

This evaluation consists in looking for a value approximating integral L_i, starting from the sum of the g function weighted by w_j in J evaluation points. The weighting coefficients corresponding to the number of evaluation points are provided in Abramovitz and Stegun (1985) for example.

In their application, Butler and Moffit show the relatively reduced stability of estimators with a number of evaluation points equal to five. In our case, it turns out that nine evaluation points give results very similar to those obtained with the LIMDEP software. However, certain estimations may require up to twenty evaluation points to stabilize the results.

- **Taking into account the problem of state dependence versus unobserved heterogeneity**

Some authors (Feller, 1943; Heckman, 1981a and b; Pinquet, 2000; Honoré and Kyriazidou, 2000; …) have looked at the question of identifying the nature of the individual heterogeneity characterizing longitudinal data whose key variable is discrete.

Heckman (1981a) provides a very instructive explanatory example for situations where an individual having been involved in a past event will have a greater chance of encountering the same event in the future than an individual who never experienced such an event. In other words, in a case involving the choice of a product such as automobile insurance coverage, an individual having made this choice in the past will tend to make the same choice in the future, a possibility not open to individuals having never made such a choice.

In statistics, this phenomenon is called state dependence. Heckman considers that state dependence arises when there is a non-uniform distribution of the dates on which the phenomenon studied occurs. He provides two explanations for this phenomenon: true state dependence and spurious state dependence. The first explanation suggests that a lag (or lags) on the dependent variable must enter the list of explanatory variables for the model one is attempting to estimate. As for the second explanation (i.e. spurious state dependence), it veers towards explanation of the temporal correlation of the residuals generated by an unobserved heterogeneity owing to a permanent, unobserved individual effect.

In our case, choice of insurance coverage may be constant over a given period of time. This is effectively the case in our sample of data: from wave to wave of the survey on the insurance-choice variable, there are few changes of state (only 10% of policy-holders switched from all-risk insurance to third-party insurance between 1995 and 1997). This points to true state dependence. However, a serial correlation of the residuals estimation of the random-effects probit model for insurance choices should not be excluded. In this case, it is difficult to come to any conclusion about the nature of the underlying dynamics. Temporal serial correlation of the residuals may, in effect, be caused by some phenomenon other than state dependence: heteroscedasticity or unobserved heterogeneity, for example.

6. Analysis of the results

We first estimated the random-effects probit model for choice of insurance coverage, using only one explanatory variable: the one related to total number of accidents occurring in 1995, 1996, and 1997. The results of this estimation show that the "total number of accidents" variable has a positive and largely significant coefficient (Table 1).

Table 1: Random-effects probit model (*incomplete panel*)
Variable explained: Choice of insurance coverage
(all-risk vs third-party insurance)

Variable	coefficient	standard deviation	T-stat	P value
Constant	0.6606	0.0203	32.4820	< 0.0000
NBACC	**0.1684**	**0.0284**	**5.9202**	**< 0.0000**
Rho	0.4844	0.0205	23.6234	< 0.0000
Log-likelihood		-9,850.95		
Number of observations		16,399		
Number of individuals		11,506		

In a second step, we introduced into the vector of the explanatory values related to the expected number of accidents: *E(NBACC)* derived from the negative binomial model estimated crosssectionally for 1995, 1996, and 1997. We have introduced also the time (*t*) variable to take into account the trend effect. That gave us the following estimations (Table 2):

Table 2: Random-effects probit model (with *E(NBACC)*)
Variable explained: Choice of insurance coverage
(all-risk vs third-party insurance)

Variable	coefficient	Standard-deviation	T-stat	P_value
Constant	1.1227	0.1368	8.2050	< 0.0000
NBACC	**0.0793**	**0.0712**	**1.1130**	**0.2658**
E(NBACC)	**8.3124**	**0.3390**	**24.5230**	**< 0.0000**
t	-0.1345	0.0410	-3.2760	0.0011
Rho	0.9644	0.0027	359.8860	< 0.0000
Log-likelihood		-8,137.47		
Number of observations		16,399		
Number of individuals		11,506		

The number of accidents does not vary significantly from zero whereas the predictive variable for the number of accidents is largely significant. This means that there is no residual asymmetric information in the data. The time variable has a negative and significant coefficient, indicating, for the period, a downward trend in the choice of all-risk coverage as compared to third-party insurance coverage. We now want to know which variables may be used to approximate the expected number of accidents effect.

In a third step, we redid the regression by introducing the *bonus-malus* variable into the list of explanatory variables. As we saw in Table 7, the *bonus-malus* variable is positively significant in explaining the number of accidents. However, it may happen that the *bonus-malus* is strongly correlated with the expected number of accidents for the reference period. In our case, the coefficient of correlation between these two variable is 0.05, which is very weak.

The *bonus-malus* variable is continuous in the original database and it is distributed from 0.5 to 3.5 with a value equal to 1 meaning that the driver has neither a bonus nor a malus. Hence, to capture its effect on the insurance coverage choice, we have constructed three classes of this variable[5]. The first one is BMINF1, for a *bonus-malus* less than 1 (i.e., to have a bonus). The second category is BMEGAL1, for a *bonus-malus* equal to 1 (i.e., neither bonus nor malus). The third category is BMSUP1, for a *bonus-malus* greater than 1 (i.e., to have a malus). The BMEGAL1 is used as the of reference category in the following regressions.

Table 3: Random-effects probit model (with *bonus-malus*)
Variable explained: Choice of insurance coverage
(all-risk vs third-party insurance)

Variable	Coefficient	standard deviation	T-stat	P_value
Constant	21.5614	3.2405	6.6537	< 0.0000
NBACC	**0.0319**	**0.0592**	**0.5385**	**0.5902**
E(NBACC)	**7.2057**	**0.2893**	**24.9064**	**< 0.0000**
BMINF1	**2.5153**	**0.2246**	**11.1972**	**< 0.0000**
BMSUP1	**0.1594**	**0.3067**	**0.5199**	**0.6031**
t	-0.2493	0.0339	-7.3605	< 0.0000
Rho	0.8916	0.0050	179.4467	< 0.0000
Log-likelihood	-8,175.75			
Number of observations	16,399			
Number of individuals	11,506			

The results of the regression confirm our prediction: the bonus category has a positive and significant parameter (Table 3): good risks choose all-risk coverage because the price is lower for them. Also, following introduction of the *bonus-malus* variable, the expected number of accidents variable remains very significant. This result shows that the *bonus-malus* variable, though significant for one category in the regression, does not add any more information about risks than do predictions of accidents *E(NBACC)*. In other words, just as with *a priori* pricing,

[5] Note however that the *bonus-malus* variable is continuous in the accident distribution estimations (Table 7). We did also consider a continuous *bonus-malus* variable in the choice of insurance equations. The conclusions are the same. Results are available from the authors.

the addition of *a posteriori* pricing has not succeeded in capturing the individual observable risk represented by the expected number of accidents.

This may perhaps be explained by the fact that recent risk has a determining influence on the decisions policyholders make in their choice of insurance coverage. In what happens, the choice of all-risk insurance would seem to result from a driver's adjusting to the risk he poses during the period.

In a fourth step, with a view to eliminating the significant sign of the "accident prediction" variable, we added explanatory variables including mileage (exposure-to-risk variable) along with other classification variables used by insurers in the regression component. That led to the estimations contained in Table 4.

Table 4: Random-effects probit model (with classification variables)
Variable explained: Choice of insurance coverage
(all-risk vs third-party insurance)

Variable	coefficients	standard deviation	T-stat	P value
Constant	-1.4694	0.3055	-4.8094	0.0000
NBACC	0.0945	0.0610	1.5475	0.1218
E(NBACC)	4.6416	0.3958	11.7256	0.0000
BMNF1	1.7472	0.2855	6.1191	0.0000
BMSUP1	-0.1369	0.3694	-0.3707	0.7109
KMU5	-0.5425	0.1085	-4.9988	0.0000
KM1015	0.6796	0.0893	7.6086	0.0000
KM1520	1.0791	0.1029	10.4905	0.0000
KM2030	0.9725	0.1197	8.1272	0.0000
KMO30	0.7731	0.1711	4.5190	0.0000
FARMER	0.0983	0.2623	0.3748	0.7078
ARTISAN	-1.3410	0.2114	-6.3432	0.0000
MANAGER	-0.5702	0.1651	-3.4530	0.0006
PROF	0.6478	0.2732	2.3715	0.0177
INTPROF	-0.5156	0.1218	-4.2338	0.0000
TEACHER	0.4611	0.2155	2.1401	0.0324
EMPLOYEE	-0.8362	0.1168	-7.1586	0.0000
WORKER	-1.5529	0.1286	-12.0778	0.0000

STUDENT	-0.9307	0.2116	-4.3991	0.0000
OTHPROF	-1.1930	0.1389	-8.5912	0.0000
AV0	4.3033	0.2171	19.8258	0.0000
AV12	3.0577	0.1153	26.5108	0.0000
AV69	-0.6897	0.0716	-9.6331	0.0000
AV10+	0.8472	0.0604	14.0259	0.0000
C2	-1.6095	0.1587	-10.1434	0.0000
C3	-1.3960	0.2323	-6.0090	0.0000
C4	-1.1864	0.3211	-3.6944	0.0002
C5	-2.6494	0.4227	-6.2676	0.0000
C2*A1820	-1.4746	3.3290	-0.4430	0.6578
C2*A2534	0.6128	0.2004	3.0577	0.0022
C2*A3544	1.2862	0.2156	5.9661	0.0000
C2*A4554	1.1906	0.3014	3.9507	0.0001
C2*A5564	1.1288	0.6287	1.7956	0.0726
C3*A1820	0.7957	1.5269	0.5211	0.6023
C3*A2534	0.3001	0.2712	1.1066	0.2685
C3*A3544	1.5788	0.2797	5.6450	0.0000
C3*A4554	1.9362	0.2898	6.6805	0.0000
C3*A5564	2.0453	0.3247	6.2989	0.0000
C3*A65+	1.7154	0.6269	2.7363	0.0062
C4*A1820	1.3466	1.6690	0.8069	0.4198
C4*A2534	0.5833	0.3758	1.5522	0.1207
C4*A3544	1.1714	0.3843	3.0485	0.0023
C4*A4554	1.7060	0.3882	4.3946	0.0000
C4*A5564	1.3974	0.4513	3.0966	0.0020
C4*A65+	3.0474	1.1782	2.5865	0.0097
C5*A1820	1.2707	0.5792	2.1940	0.0283
C5*A2534	-0.2026	0.7671	-0.2641	0.7917
C5*A3544	-0.0519	1.8917	-0.0275	0.9781
C5*A4554	2.9153	1.4861	1.9618	0.0498
RHO	0.8820	0.0062	142.9556	0.0000
Log-likelihood			-6728.06	
Number of observations			16,399	
Number of individuals			11,506	

These new variables are the most significant in explaining the choice of all-risk insurance coverage, thus indicating that the choice of all-risk insurance is more sensitive to the condition of the vehicle and the mobility of the drivers.

The "*expected number of accidents*" variable remains significant despite the introduction of classification variables as well as age of vehicle and mileage (exposure-to-risk variable). This may be interpreted in two ways. The first hinges on comparison with the cross-sectional modeling: perhaps there are still non-linearities linked to insurance pricing which have not been accounted for and so these non-linearities must be eliminated by finding the right cross-matches between the explanatory variables and the classification variables. The second is connected with the hypothesis that there may exist a state dependence or some unobserved heterogeneity in the phenomena modeled. Since the residuals in our estimated models are serially correlated, we can draw no conclusion concerning the nature of the dynamics governing the choice of insurance coverage.[6]

So, we dated explanatory valuables which do not "move" with time (such as the SPC for example) and re-estimated the model. The results (Table 5) show that the expected number of accidents has become non-significant, whereas the *bonus-malus* categories remain significantly different than zero. In what happens, it would seem that the fact of dating the variables which do not vary over time was a way of proving the existence of a permanent unobserved individual effect. This thus excludes the hypothesis of state dependence initially envisioned and points rather towards the hypothesis of unobserved heterogeneity.

Consequently, we come to the same conclusion as with *cross-sectional* analysis (DGV, 2001): the policy holder does not have any more information than his insurer and the risk classification has succeeded in capturing individual risks.

As regards the *bonus-malus* variable, the estimated coefficient of the BMINF1 category (to have a bonus) is still positive and significant. This means that drivers who have a bonus buy more all-

[6] See Section 5 for the notion of state dependence.

risk coverage than other drivers, *ceteris paribus*. This is because of a price effect: the cost to buy more coverage when one has a bonus is less expensive. Here, the equation estimated is an equation on the demand for all-risk insurance coverage and the *bonus-malus* appears as a pricing variable: policyholders with low *bonus-malus* pay a low price and thus ask for more all-risk insurance coverage. But an important question is the following: is the presence of the *bonus-malus* variable necessary to obtain our conclusion on the residual presence of asymmetrical information in the data? Can we obtain the same result by using other classification variables?

Table 5: Random-effects probit model (*with dated variables*)
Variable explained: Choice of insurance coverage
(all-risk vs third-party insurance)

Variable	coefficients	standard deviation	T-stat	P value
Constant	-4.2313	0.7086	-5.9716	< 0.0000
NBACC	**0.0780**	**0.0660**	**1.1823**	**0.2371**
E(NBACC)	**0.1107**	**0.5294**	**0.2091**	**0.8344**
BMINF1	**1.2326**	**0.2878**	**4.2830**	**0.0000**
BMSUP1	**0.1036**	**0.3762**	**0.2753**	**0.7831**
KMU5	-0.3959	0.1167	-3.3919	0.0007
KM1015	0.3267	0.0955	3.4209	0.0006
KM1520	0.6071	0.1144	5.3093	0.0000
KM2030	0.5412	0.1294	4.1824	0.0000
KMO30	0.5288	0.1891	2.7955	0.0052
FARMER*T1	0.0180	0.0047	3.8008	0.0001
ARTISAN*T1	-0.0011	0.0038	-0.2945	0.7684
MANAGER*T1	0.0086	0.0028	3.0363	0.0024
PROF*T1	0.0115	0.0048	2.4171	0.0157
INTPROF*T1	0.0044	0.0022	1.9985	0.0457
TEACHER*T1	0.0065	0.0041	1.6071	0.1081
EMPLOYEE*T1	-0.0075	0.0022	-3.4383	0.0006
STUDENT*T1	0.0028	0.0041	0.6924	0.4887
OTHPROF*T1	-0.0024	0.0024	-0.9863	0.3240
FARMER*T2	0.0038	0.0088	0.4321	0.6657
ARTISAN*T2	-0.0170	0.0084	-2.0329	0.0421
MANAGER*T2	-0.0058	0.0076	-0.7613	0.4465

Variable	coefficients	standard deviation	T-stat	P value
PROF*T2	0.0126	0.0047	2.6621	0.0078
INTPROF*T2	-0.0052	0.0074	-0.7027	0.4823
TEACHER*T2	0.0076	0.0034	2.2190	0.0265
EMPLOYEE*T2	-0.0112	0.0073	-1.5393	0.1238
WORKER*T2	-0.0128	0.0074	-1.7289	0.0838
STUDENT*T2	-0.0100	0.0072	-1.3780	0.1682
OTHPROF*T2	-0.0082	0.0073	-1.1147	0.2650
FARMER*T3	0.0379	0.0101	3.7364	0.0002
ARTISAN*T3	0.0232	0.0091	2.5520	0.0107
MANAGER*T3	0.0267	0.0086	3.1006	0.0019
PROF*T3	0.0133	0.0043	3.0819	0.0021
INTPROF*T3	0.0343	0.0084	4.0826	0.0000
TEACHER*T3	0.0087	0.0038	2.2965	0.0217
EMPLOYEE*T3	0.0280	0.0082	3.4000	0.0007
WORKER*T3	0.0228	0.0083	2.7378	0.0062
STUDENT*T3	0.0301	0.0079	3.8233	0.0001
OTHPROF*T3	0.0243	0.0084	2.9118	0.0036
AV0*T1	0.0279	0.0032	8.6851	0.0000
AV12*T1	0.0210	0.0024	8.6309	0.0000
AV69*T1	-0.0026	0.0012	-2.1180	0.0342
AV10+*T1	0.0390	0.0017	23.4036	0.0000
AV0*T2	0.0427	0.0065	6.5232	0.0000
AV12*T2	0.0200	0.0022	8.9641	0.0000
AV69*T2	-0.0082	0.0013	-6.3824	0.0000
AV10+*T2	0.0385	0.0016	23.3970	0.0000
AV0*T3	0.0274	0.0037	7.3274	0.0000
AV12*T3	0.0197	0.0022	9.0224	0.0000
AV69*T3	-0.0252	0.0015	-16.8129	0.0000
AV10+*T3	-0.0503	0.0023	-22.1447	0.0000
C2*T1	0.0245	0.0068	3.6325	0.0003
C3*T1	0.0001	0.0068	0.0179	0.9857
C4*T1	0.0042	0.0066	0.6402	0.5220
C5*T1	0.0096	0.0073	1.3191	0.1871
C2*T2	0.0238	0.0068	3.4815	0.0005

Variable	coefficients	standard deviation	T-stat	P value
C3*T2	0.0088	0.0066	1.3294	0.1837
C4*T2	0.0117	0.0065	1.7955	0.0726
C5*T2	0.0141	0.0071	1.9943	0.0461
C2*T3	0.0624	0.0069	9.0348	0.0000
C3*T3	0.0114	0.0074	1.5466	0.1220
C4*T3	0.0151	0.0074	2.0346	0.0419
C5*T3	0.0180	0.0079	2.2948	0.0218
C2*A1820	-2.8612	4.1199	-0.6945	0.4874
C2*A2534	0.5787	0.2238	2.5859	0.0097
C2*A3544	1.1483	0.2366	4.8530	0.0000
C2*A4554	1.0972	0.2916	3.7625	0.0002
C2*A5564	0.5293	0.7081	0.7475	0.4548
C3*A1820	2.0836	2.0760	1.0036	0.3156
C3*A2534	0.0560	0.2817	0.1987	0.8425
C3*A3544	1.1619	0.2921	3.9782	0.0001
C3*A4554	1.4632	0.3047	4.8017	0.0000
C3*A5564	1.7559	0.3515	4.9949	0.0000
C3*A65+	1.3918	0.6153	2.2620	0.0237
C4*A1820	2.8255	3.3711	0.8381	0.4020
C4*A2534	0.3133	0.3991	0.7850	0.4325
C4*A3544	0.7024	0.4033	1.7419	0.0816
C4*A4554	1.2803	0.4163	3.0753	0.0021
C4*A5564	0.9134	0.4567	1.9998	0.0455
C4*A65+	2.2031	1.5140	1.4551	0.1457
C5*A1820	1.3738	0.5644	2.4340	0.0149
C5*A2534	-0.5752	0.8046	-0.7149	0.4747
C5*A3544	-0.2897	1.6609	-0.1744	0.8615
C5*A4554	2.7617	2.2734	1.2148	0.2245
Rho	0.8779	0.0076	115.1477	0.0000
Log-likelihood		-5,846.23		
Number of observations		16,399		
Number of individuals		11,506		

Table 6 shows that the absence of the *bonus-malus* variable does not affect our conclusion about the presence of residual asymmetric information in the data. Note that the corresponding accident distribution estimations do not contain also the *bonus-malus* variable (detailed results are available from the authors upon request).

Table 6: Random-effects probit model (without *bonus-malus* variable)
Variable explained: Choice of insurance coverage
(all-risk vs third-party insurance)

Variable	**coefficients**	**standard deviation**	**T-stat**	**P value**
CONSTANT	-2.9052	0.6850	-4.2414	< 0.0000
NBACC	**0.0882**	**0.0654**	**1.3486**	**0.1775**
E(NBACC)	**-0.6102**	**0.4759**	**-1.2822**	**0.1998**
KMU5	-0.4126	0.1172	-3.5195	0.0004
KM1015	0.3442	0.0960	3.5865	0.0003
KM1520	0.6582	0.1146	5.7437	< 0.0000
KM2030	0.5846	0.1292	4.5263	< 0.0000
KMO30	0.5839	0.1862	3.1367	0.0017
FARMER*T1	0.0181	0.0047	3.8295	0.0001
ARTISAN*T1	-0.0010	0.0038	-0.2542	0.7994
MANAGER*T1	0.0095	0.0028	3.3446	0.0008
PROF*T1	0.0114	0.0048	2.3821	0.0172
INTPROF*T1	0.0051	0.0022	2.3397	0.0193
TEACHER*T1	0.0073	0.0042	1.7542	0.0794
EMPLOYEE*T1	-0.0078	0.0022	-3.5701	0.0004
STUDENT*T1	0.0019	0.0042	0.4453	0.6561
OTHPROF*T1	-0.0020	0.0024	-0.8518	0.3943
FARMER*T2	0.0052	0.0093	0.5548	0.5791
ARTISAN*T2	-0.0154	0.0088	-1.7529	0.0797
MANAGER*T2	-0.0034	0.0081	-0.4253	0.6706
PROF*T2	0.0133	0.0047	2.8260	0.0047
INTPROF*T2	-0.0026	0.0079	-0.3358	0.7370
TEACHER*T2	0.0082	0.0035	2.3398	0.0193
EMPLOYEE*T2	-0.0093	0.0078	-1.1887	0.2346
WORKER*T2	-0.0108	0.0079	-1.3613	0.1734

STUDENT*T2	-0.0091	0.0077	-1.1896	0.2342
OTHPROF*T2	-0.0060	0.0078	-0.7718	0.4402
FARMER*T3	0.0367	0.0104	3.5359	0.0004
ARTISAN*T3	0.0223	0.0093	2.3888	0.0169
MANAGER*T3	0.0268	0.0089	3.0305	0.0024
PROF*T3	0.0124	0.0043	2.8652	0.0042
INTPROF*T3	0.0342	0.0087	3.9362	0.0001
TEACHER*T3	0.0094	0.0037	2.5594	0.0105
EMPLOYEE*T3	0.0271	0.0085	3.1911	0.0014
WORKER*T3	0.0216	0.0086	2.5231	0.0116
STUDENT*T3	0.0297	0.0081	3.6477	0.0003
OTHPROF*T3	0.0235	0.0086	2.7257	0.0064
AV0*T1	0.0279	0.0033	8.3689	< 0.0000
AV12*T1	0.0210	0.0025	8.5689	< 0.0000
AV69*T1	-0.0027	0.0012	-2.2231	0.0262
AV10+*T1	0.0396	0.0017	23.7297	< 0.0000
AV0*T2	0.0405	0.0060	6.7731	< 0.0000
AV12*T2	0.0202	0.0022	8.9630	< 0.0000
AV69*T2	-0.0082	0.0013	-6.3556	< 0.0000
AV10+*T2	0.0391	0.0017	23.6388	< 0.0000
AV0*T3	0.0274	0.0038	7.2395	< 0.0000
AV12*T3	0.0196	0.0022	8.9714	< 0.0000
AV69*T3	-0.0255	0.0015	-16.9573	< 0.0000
AV10+*T3	-0.0511	0.0023	-22.3592	< 0.0000
C2*T1	0.0241	0.0071	3.3937	0.0007
C3*T1	-0.0009	0.0071	-0.1298	0.8967
C4*T1	0.0031	0.0070	0.4430	0.6578
C5*T1	0.0089	0.0076	1.1730	0.2408
C2*T2	0.0234	0.0072	3.2581	0.0011
C3*T2	0.0061	0.0066	0.9283	0.3533
C4*T2	0.0093	0.0066	1.4195	0.1558
C5*T2	0.0119	0.0072	1.6653	0.0959
C2*T3	0.0626	0.0072	8.6463	< 0.0000
C3*T3	0.0121	0.0074	1.6287	0.1034
C4*T3	0.0159	0.0075	2.1225	0.0338

C5*T3	0.0190	0.0079	2.4015	0.0163
C2*A1820	-5.0170	1.9147	-2.6202	0.0088
C2*A2534	0.5890	0.2249	2.6190	0.0088
C2*A3544	1.1561	0.2373	4.8713	< 0.0000
C2*A4554	1.1058	0.2913	3.7957	0.0001
C2*A5564	0.4627	0.6922	0.6685	0.5038
C3*A1820	2.1877	2.1610	1.0124	0.3114
C3*A2534	0.0526	0.2861	0.1838	0.8541
C3*A3544	1.1782	0.2953	3.9894	0.0001
C3*A4554	1.4680	0.3081	4.7650	< 0.0000
C3*A5564	1.7905	0.3522	5.0838	< 0.0000
C3*A65+	1.3435	0.5999	2.2396	0.0251
C4*A1820	2.9190	3.1833	0.9170	0.3592
C4*A2534	0.2721	0.3983	0.6831	0.4945
C4*A3544	0.7305	0.4022	1.8162	0.0694
C4*A4554	1.2686	0.4163	3.0474	0.0023
C4*A5564	0.9267	0.4568	2.0286	0.0425
C4*A65+	2.2220	1.5368	1.4459	0.1482
C5*A1820	1.1847	0.5684	2.0842	0.0372
C5*A2534	-0.7854	0.8285	-0.9480	0.3432
C5*A3544	0.1018	1.9551	0.0521	0.9585
C5*A4554	1.2666	1.6468	0.7691	0.4418
RHO	0.8793	0.0076	116.4324	< 0.0000
Log-likelihood		-5860.94		
Number of observations		16,399		
Number of individuals		11,506		

Conclusion

In this analysis, we have attempted to extend the Dionne, Gouriéroux, and Vanasse test (2001) to the case of individual longitudinal data. Our main objective was to find out if the conclusions from successive tests conducted on cross-sectional studies would hold true in the case of panel data. Our panel was constructed over the period 1995–1997.

The results of the different regressions allowed us to come to the same conclusion as in the case of the cross-sectional tests: absence of any informational advantage of the policy-holder over his insurer and efficiency of risk classification in accounting for individual risks. We also show that the *bonus-malus* is a pricing variable (in the equation on choice of coverage) as well as a signal of type of risk (in the equation on accidents). It thus constitutes a significant element in the pricing mechanism, making it possible to introduce incentives for drivers with regard to the prevention of traffic accidents.

However, its absence does not affect the conclusion about the presence of residual asymmetric information in the data. It seems that its overall effect can be replicated by other appropriate risk classification variables. Finally, since the test shows that there is no residual asymmetric information between insurers and insureds in this market, there is non need to add a further step of analysis to separate moral hazard from adverse selection as in Dionne and Gagné (2002).

References

M. Abramovitz, and I. Stegun [1985], Handbook of Mathematical Functions, *Dover Press*, New-York.

J. S. Butler and R. Moffit [1982], A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit, *Econometrica*, 50, 761-764.

Conseil des Communautés Européennes [1992], directive No 92/49/CEE dite "Troisième directive assurance non-vie", *JOCE*, Commission européenne, Bruxelles.

S. J. Chandler [1999], Insurance Regulation, mimeo, University of Houston Law Centre.

P-A. Chiappori [2000], Econometric Models of Insurance Under Asymmetric Information, In G. Dionne (ed.), Handbook of Insurance, *Kluwer Academic Publishers*, 363-392.

P-A. Chiappori, and B. Salanié [1997], Empirical Contracts Theory: The Case of Insurance Data, *European Economic Review*, 41, 943-951.

P. A. Chiappori, and B. Salanié [2000], Testing for Asymmetric Information in Insurance Markets, *Journal of Political Economy*, 108, 1, 56-78.

P. A. Chiappori, and B. Salanié [2002], Testing Contracts Theory: a Survey of Some Recent Work. Mimeo, CREST, France.

R. Cooper, and B. Hayes [1987], Multi-period Insurance Contracts, *International Journal of Industrial Organization*, 5, 211-231.

K. Crocker, and A. Snow [1986], The Efficiency Effects of Categorical Discrimination in the Insurance Industry, *Journal of Political Economy*, 94, 321-344.

M. Dahchour [2002], Tarification de l’assurance automobile, utilisation du permis à points et incitation à la sécurité routière : une analyse empirique, Ph. D Dissertation, University of Paris X-Nanterre, France.

M. Dahchour, and S. Lassarre [2001], Road Risk Analysis using Panel Data of French Drivers, In G. Govaert, J. Janssen, N. Limnios (Eds.). Proceedings of the 10th International Symposium on Applied Stochastic Models and Data Analysis (ASMDA), vol 1. 348-353. Université de Technologie Compiègne, France.

G. Dionne [2002], Insurance Regulation in Other Industrial Countries, In J. D. Cummins (ed.) Deregulating Property-Liability Insurance; AEI-Brookings Joint Center For Regulatory Studies, Washington, 341-396.

G. Dionne, and N. Doherty [1994], Adverse Selection, Commitment and Renegotiation: Extension to and Evidence from Insurance Markets, *Journal of Political Economy*,102, 209-235.

G. Dionne, N. Doherty, and N. Fombaron [2000], Adverse Selection in Insurance Markets, in Handbook of Insurance, G. Dionne (Ed.), *Kluwer Academic Publishers*, 185-243.

G. Dionne et R. Gagné [2002], Replacement Cost Endorsement and Opportunistic Fraud in Automobile Insurance, WP No 00-01, Risk Management Chair, HEC-Montreal, Forthcoming *Journal of Risk and Uncertainty*.

G. Dionne, C. Gouriéroux and C. Vanasse [1998], Evidence of Adverse Selection in Automobile Insurance Markets. In Automobile Insurance: Road Safety, New Drivers, Risks, Insurance Fraud and Regulation, G. Dionne and C. Laberge-Nadeau (Eds.), *Kluwer Academic Publishers*, 13-46.

G. Dionne, C. Gouriéroux, and C. Vanasse [2001], Testing for Evidence of Adverse Selection in the Automobile Insurance Market: A Comment. *Journal of Political Economy*, 109, 444-453.

A. Finkelstein et J. Poterba [2000], Adverse Selection in Insurance Markets: Policyholder Evidence from the U.K. Annuity Market, *NBER* Working Paper No. 8045.

N. Fombaron [1997], Contrats d'assurance dynamiques en présence d'antisélection : les effets d'engagement sur des marchés concurrentiels, Ph.D Dissertation, University Paris X-Nanterre. France.

N. Fombaron [2002], Dérégulation du bonus-malus en assurance automobile et incitations à la sécurité routière : le cas de la France, *Assurances*, 69, 4, 589-601.

W. Greene [1998], Limdep, Version 7.0, User's Manual, Econometric Software, New York.

W. Greene [2000], Econometric Analysis, Fourth Edition, Prentice-Hall, New Jersey.

I. Hendel, and A. Lizzeri (2000), The Role of Commitment in Dynamic Contracts: Evidence from Life Insurance, Mimeo, Princeton University.

D. K. Guilkey, and J. L. Murphy [1993], Estimation and testing in the random effects probit model, *Journal of Econometrics*, 59. 301-317.

J. J. Heckman [1981 a], Panel data, in C. F. Manski, and D. MacFadden (eds.), Structural Analysis of Discrete Data with Econometric Applications, Chap 3, *MIT Press*.

J. J. Heckman [1981 b], Heterogeneity and State Dependence, in S. Rosen (ed.), Studies in Labor Markets, *Chicago University Press*. 91-139.

B. E. Honoré, and E. Kyriazidou [2000], Panel Data Discrete Choice Models with Lagged Dependent Variables. *Econometrica.* Vol. 68, No. 4, 939-874.

C. Hsiao [1996], Logit and Probit models, in Mátyàs, and Sevestre (Eds.), The Econometrics of Panel Data: Handbook of Theory with Applications, *Kluwer Academic Publishers*. Dordrecht.

H. Kunreuther, and M. Pauly [1985], Market Equilibrium with Private Knowledge, *Journal of Public Economics*, 26, 269-288.

K.M. Murphy, and R.H. Topel [1985], Estimation and Inference in Two-Step Econometric Models. *Journal of Business and Economic Statistics*, 3, 370-379.

T. Nielssen [2000], Consumer Lock-in with Asymmetric Information, *International Journal of Industrial Organization*, 18, 649-666.

P. Picard [2000], Les nouveaux enjeux de la régulation des marches d'assurance, WP No 2000-53, *Thema*, Université Paris X Nanterre.

J. Pinquet [1999], Une analyse des systèmes bonus-malus en assurance automobile, *Assurances*, 67, 2, 241-249.

J. Pinquet [2000], Experience Rating through Heterogeneous Models. In G. Dionne (Ed.) Handbook of Insurance, *Kluwer Academic Publishers*.

R. Puelz, and A. Snow [1994], Evidence on Adverse Selection : Equilibrium Signaling and Cross Subsidization on the Insurance Market, *Journal of Political Economy*, 102, 236-257.

D.Richaudeau [1998], Le marché de l'assurance automobile en France, *Assurances*, 66, 3, 423-458.

D. Richaudeau [1999], Automobile Insurance Contracts and Risk of Accident : an Empirical Test Using French Individual Data, *Geneva Papers on Risk and Insurance Theory*, 24, 97-114.

G. Rosenwald [2000], Devenir du bonus-malus, Communication at « La société Française de Statistique », November 29, 2000, Paris.

M. Rothschild, and J. Stiglitz [1976], Equilibrium in Insurance Markets: An Essay on the Economics of Imperfect Information, *Quarterly Journal of Economics*, 90, 630-649.

C. Wilson [1977], A Model of Insurance Markets with Incomplete Information, *Journal of Economic Theory*, 16, 167-207.

Appendix

List of explanatory variables

Age and sex of principal driver
SEXM = 1 if the driver is male, otherwise 0. A1820 = 1 if the age is between 18 and 20, otherwise 0. A2124 = 1 the age is between 21 and 24, otherwise 0. (reference) A2534 = 1 if the age is between 25 and 34, otherwise 0. A3544 = 1 is the age is between 35 and 44, otherwise 0. A4554 = 1 if age is between 45 and 54, otherwise 0. A5564 = 1 if age is between 55 and 64, otherwise 0. A65+ = 1 if age is equal to 65 and over, otherwise 0.
Socio-professional category of driver
FARMER = 1 if driver is a farmer, otherwise 0. ARTISAN = 1 if driver is an artisan, otherwise 0. MANAGER = 1 if driver is a manager, otherwise 0. PROF = 1 if the driver is a professor or teacher, otherwise 0. STUDENT = 1 if driver is a student, otherwise 0. INTERPTOF = 1 if driver is of intermediate profession, otherwise 0. EMPLOYEE = 1 if driver is an employee, otherwise 0. WORKER = 1 if driver is a worker, otherwise 0. OTHPROF = 1 if driver is of some other profession, otherwise 0. RETRIRED = 1 if driver is retired, otherwise 0. (reference)
Driving experience
EXP01 = 1 if the license is between 0 and 1 year, otherwise 0. EXP23 = 1 if license is between 2 and 3 years, otherwise 0. (reference) Exp 410 = 1 if license is between 4 and 10 years , otherwise 0. EXP 11+ = 1 if the license is 11 years or more, otherwise 0.
Driver's *bonus-malus* coefficient
BMINF1=1 if the *bonus-malus* coefficient is less than 1, otherwise 0.
BMEGAL1=1 if the *bonus-malus* coefficient is equal to 1, otherwise 0. (reference)
BMSUP1=1 if the *bonus-malus* coefficient is greater than 1, otherwise 0.
Household income
HI70 = 1 if the household to which the driver belongs has an annual income less than 70, 000 F, otherwise 0. HI70to100 = 1 if the household to which the driver belongs has an annual income between 70 00F and 100 000 F, otherwise 0. (reference)

HI100to130 = 1 if the household to which the driver belongs has an annual income between 100 000F and 130 000F, otherwise 0. HI130to160 = 1 if the household to which the driver belongs has an annual income between 130 000F and 160 000F, otherwise 0. HI160 to 190 = 1 if the household to which the driver belongs has an annual income between 160 000F and 190 000F, otherwise 0. HI190 to230 = 1 if the household to which the driver belongs has an annual income between 190 000 and 230 000, otherwise 0. HI230to290 = 1 if the household to which the driver belongs has an annual income between 230 000 F and 290 000F, otherwise 0.
Region of residence
PARIS = 1 if the driver resides in the Parisian region, otherwise 0. (reference) NORTH = 1 if the driver resides in the North, otherwise 0. EAST = 1 if the driver resides in the East, otherwise 0. WEST = 1 if the driver resides in the West, otherwise 0. SOUTHWEST = 1 if the driver resides in the Southwest, otherwise 0. SOUTHEAST = 1 if the driver resides in the Southeast, otherwise 0. MEDITERA = 1 if the driver resides in the Mediterranean region, otherwise 0.
Network used
CITY = 1 if the network used is mainly urban, otherwise 0. ROAD = 1 if the network used is mainly country roads, otherwise 0. (reference) HIGHW = 1 if the network used is mainly highway, otherwise 0.
Age of vehicle
AV0 = 1 if the vehicle is new (year of purchase corresponding to year of survey), otherwise 0. AV12 = 1 if the vehicle is 1 or 2 years old, otherwise 0. AV35 - 1 if the vehicle is 3, 4, or 5 years old, otherwise 0. (reference) AV69 = 1 if the vehicle is 6,7, 8, or 9 years old, otherwise 0. AV10+ = 1 if the vehicle is 10 year old or over, otherwise 0.
Group of vehicle
GROU7U = 1 if the vehicle belongs to group 7 or under, otherwise 0. GROU89 = 1 if the vehicle belongs to groups 8 or 9, otherwise 0. GROU1011 = 1 if the vehicle belongs to groups 10 or 11, otherwise 0. (reference) GROU1213 = 1 if the vehicle belongs to groups 12 or 13, otherwise 0. GROU140 = 1 if the vehicle belongs to the group 14 or over, otherwise 0.
Use of vehicle
DAILY = 1 is the vehicle is used daily, otherwise 0. (reference) ALMDAIL = 1 if the vehicle is used almost daily, otherwise 0. LESSOFT = 1 if the vehicle is used less often, otherwise 0.

WEEKEND = 1 if the vehicle is used only on weekends, otherwise 0. ALMNEV = 1 if the vehicle is almost never used, otherwise 0.
NONPROP=1 if the vehicle is owned by the driver, otherwise 0.
DIESEL=1 if the vehicle uses gas-oil as a fuel, otherwise 0.
Mileage
KMU5 = 1 if the mileage is under 5 000 km a year, otherwise 0. KM510 = 1 if the mileage is between 5 000 and 10 000 km a year, otherwise 0. KM1015 = 1 if the mileage is between 10 000 and 15 000 km a year, otherwise 0. (reference) KM1520 = 1 if the mileage is between 15 000 and 20 000 km a year, otherwise 0. KM2030 = 1 if the mileage is between 20 000 and 30 000 km a year, otherwise 0. KMO30 = 1 if the mileage is over 30 000 km a year, otherwise 0.
Occasional users
NOOCC = 1 if there is no occasional user of the vehicle, otherwise 0. (reference) YOCCL = 1 if the occasional user is young (male or female under 25) having driven less than 10% of total mileage, otherwise 0. YOCCM = 1 if the occasional user is young and has driven more than 10% of the total mileage, otherwise 0. MPCCM = 1 if the occasional user is man over 25 having driven less than 10% of the total mileage, otherwise 0. MOCCM = 1 if the occasional user is a man over 25 having driven more than 10% of the total mileage, otherwise 0. WOCCL = 1 if the occasional user is a woman over 25 having driven less than 10% of the total mileage, otherwise 0. WOCCM = 1 if the occasional user is a woman over 25 having driven more than 10% of the total mileage, otherwise 0.

Table 7: Estimation of negative binomial model period by period over 1995-1997
(Dependant variable: total number of accidents)

Variable	1995		1996		1997	
	Coefficient	**T-stat**	**Coefficient**	**T-stat**	**Coefficient**	**T-stat**
CONSTANT	-2.7736	-8.0426	-2.6424	-9.1161	-2.3902	-7.5116
(Age and sex of driver)						
Sexm	-0.1524	-1.5025	-0.0648	-0.7648	-0.2714	-3.0055
A1820	-0.3027	-0.6198	-0.5287	-1.0928	0.7039	1.5276
A2124	0.2016	1.0010	0.1198	0.6109	0.0947	0.3881
A2534			Reference			
A3544	0.0138	0.1026	0.0434	0.3777	-0.0099	-0.0820
A4554	-0.1273	-0.7048	-0.0911	-0.5821	-0.0541	-0.3307
A5564	-0.1040	-0.4198	-0.3378	-1.5535	0.0699	0.3112
A65+	0.2626	0.8604	-0.1834	-0.6714	0.0395	0.1392
(Driving experience)						
EXP01	0.8287	1.5335	1.2237	2.6142	-0.9238	-1.3398
EXP12	0.6018	1.2247	0.9019	2.0426	-0.6629	-1.2988
EXP34			Reference			
EXP510	-0.1716	-0.2684	0.6509	1.0103	-0.6822	-0.9054
EXP11+	-0.3170	-2.0306	-0.0222	-0.1585	-0.2045	-1.3633
***Bonus-malus* (continuous)**	0.6874	2.7907	1.2930	6.7611	0.7705	3.2279
(Driver's SPC)						
RETIRED			Reference			
FARMER	-0.4802	-1.1623	-0.6992	-1.6345	-0.6553	-1.3431
ARTISAN	-0.5591	-1.6492	0.0484	0.1892	-0.0308	-0.1213
MANAGER	-0.1353	-0.5505	0.0063	0.0310	0.2153	1.0940
PROF	0.1840	0.6689	0.1756	0.8490	-0.2152	-0.9061
INTPROF	-0.0891	-0.4119	-0.1138	-0.6211	-0.0128	-0.0701
TEACHER	-0.0677	-0.3189	0.2335	1.4486	0.2475	1.3758
EMPLOYEE	-0.4560	-2.0324	-0.2156	-1.1508	-0.1286	-0.6848
WORKER	-0.4944	-2.1119	-0.3543	-1.8253	-0.105	-0.5469
STUDENT	-0.0204	-0.0670	-0.4195	-1.4410	-0.0132	-0.0404
OTHPROF	-0.1761	-0.7526	-0.1143	-0.5568	-0.1475	-0.7153
(Region)						
PARIS			Reference			
NORTH	-0.2471	-1.5762	-0.0004	-0.0035	-0.3371	-2.2280
EAST	-0.1086	-0.7442	-0.1155	-0.9285	-0.0233	-0.1861

Variable	1995		1996		1997	
	Coefficient	**T-stat**	**Coefficient**	**T-stat**	**Coefficient**	**T-stat**
WEST	-0.0117	-0.0873	-0.0620	-0.5631	-0.0314	-0.2762
SOUTHWEST	0.0634	0.4608	0.0440	0.3724	-0.2244	-1.7355
SOUTHEAST	0.0518	0.4194	0.0761	0.7237	-0.2245	-1.8600
MEDITERA	-0.0351	-0.2534	0.0354	0.3114	0.0052	0.0446
(Vehicle: group and age)						
GROU7U	-0.1355	-1.5542	-0.1348	-1.3656	-0.1142	-1.1996
GROU89	-0.0902	-0.8650	0.0332	0.3158	-0.0404	-0.3004
GROU1011			Reference			
GROU1213	-0.1879	-1.8738	0.1365	1.4658	-0.0525	-0.4561
GROU14P	0.3462	2.4735	0.2376	1.9961	-0.133	-0.8428
AV0	-0.0428	-0.2917	-0.1226	-0.9487	-0.2206	-1.4895
AV12	-0.0471	-0.4349	0.0341	0.3893	0.0888	0.9427
AV35			Reference			
AV69	-0.0460	-0.4753	0.0554	0.6530	0.0539	0.6160
AV10PLUS	0.4953	4.5781	0.2304	2.4502	0.3671	3.2033
NONPROP	0.6325	2.7844	0.1533	0.6229	0.3422	1.3760
DIESEL	-0.0191	-0.2010	0.0608	0.7727	0.0218	0.2636
(Mileage)						
KMU5	-0.8541	-0.7780	-2.9919	-2.8837	-0.6385	-0.5944
KM510			Reference			
KM1015	0.3116	2.5054	0.0111	0.1089	0.1796	1.5894
KM1620	0.4877	3.6258	0.1793	1.6077	0.3101	2.5089
KM2030	0.4965	3.4519	0.2425	2.0746	0.5791	4.6439
KM030	0.8817	4.9679	0.5079	3.4769	0.5646	3.4033
(Network)						
CITY	0.2725	2.6999	0.2159	2.5963	0.2407	2.6451
HIGHWAY	0.1176	1.3772	0.1276	1.7938	0.2553	3.3070
ROAD			Reference			
(Use)						
DAILY			Reference			
ALMDAI	0.1399	1.4527	0.0192	0.2338	-0.0693	-0.7813
WEEKEND	-0.6399	-1.9227	0.1615	0.9324	-0.3221	-1.3152
ALMNEV	-0.4632	-1.0408	-0.3707	-0.9764	-0.0014	-0.0036
LESSOFT	-0.0100	-0.0790	-0.1263	-1.1918	-0.0582	-0.5132
(Income)						

Variable	**1995**		**1996**		**1997**	
	Coefficient	**T-stat**	**Coefficient**	**T-stat**	**Coefficient**	**T-stat**
HI70	0.0567	0.1883	0.2584	0.3872	-0.0195	-0.0791
HI70to100			Reference			
HI100to130	0.1304	0.3713	-0.5332	-1.0993	-0.1919	-1.2534
HI130to160	-0.3075	-1.1698	-0.2234	-1.0361	-0.2482	-1.8385
HI160to190	-0.0305	-0.1948	-0.0918	-0.7139	-0.1927	-1.4543
HI190to230	-0.1277	-0.9580	-0.0834	-0.7590	-0.2099	-1.4699
HI230to290	-0.0571	-0.5556	-0.0886	-1.0442	-0.1943	-1.8091
(Occasional users)						
NOOCC			Reference			
WOCCL	0.3149	2.5705	-0.0513	-0.5031	0.0839	0.7711
MOCCL	0.3050	2.2373	-0.1132	-0.9313	-0.2502	-1.8424
WOCCM	0.4033	3.3741	0.0272	0.2563	0.1036	0.9291
MOCCM	0.2971	0.8736	-0.1555	-0.5181	0.0805	0.2625
YOCCL	0.0336	0.0652	-0.6620	-1.2559	0.8225	2.1992
YOCCM	0.7401	1.7069	0.0998	0.2392	-0.0139	-0.0227
Dispersion coefficient	**1.6600**	**6.1256**	**0.8800**	**4.6243**	**0.6263**	**4.3828**
Number of observations	5,703		5,837		5,279	
Log-likelihood	-2,468.29		-3,010.88		-2,431.44	