

Modélisation multivariée des comportements à risque des conducteurs d'automobile

Mériem Maatig

► **To cite this version:**

Mériem Maatig. Modélisation multivariée des comportements à risque des conducteurs d'automobile. 42èmes Journées de Statistique, 2010, Marseille, France, France. 2010. <inria-00494711>

HAL Id: inria-00494711

<https://hal.inria.fr/inria-00494711>

Submitted on 24 Jun 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modélisation multivariée des comportements à risque des conducteurs d'automobile

Mériem Mâatig¹

Résumé

L'objet de cette étude est d'avoir une spécification économétrique qui modélise simultanément les comportements à risque des conducteurs. En effet, l'existence d'un biais d'endogénéité peut nous amener à retenir, à tort, des variables explicatives dans une équation séparée. Pour cela nous nous appuyons sur des données collectées à partir d'une enquête. Nous avons réalisé l'enquête par questionnaire sur un échantillon représentatif de la région parisienne en termes de sexe et âge. Un modèle probit trivarié a été effectué pour tester l'existence de liens de causalité entre les trois comportements à risques des conducteurs. Dans cette modélisation, les caractéristiques qui expliquent ces trois comportements sont corrélées. Dans ce cas, l'estimation autonome de l'une de trois équations peut comporter un biais d'endogénéité. Les caractéristiques individuelles qui expliquent l'utilisation des places de parking réservées aux handicapés et la conduite sous l'emprise de l'alcool, expliquent aussi négativement la probabilité de l'utilisation du téléphone portable au volant.

Mots clés : Assurance automobile, comportements à risque, enquête, modèle probit trivarié.

Classification JEL: C1, C31, C83, G22.

Abstract

The aim of this study is to find an econometric specification that simultaneously models the risky behavior of drivers. Indeed, the existence of an endogenous bias can mislead us into retaining explanatory variables in a separate equation. The data used for this research were obtained from a questionnaire-based survey of a representative sample of the Paris population with regard to age and gender. We use trivariate probit model to test the existence of causal links between the three risky behaviors of drivers. In this modelling, the characteristics that explain these three behaviors are correlated. In this case, the autonomous estimation of one of the three equations can include an endogenous bias. The individual characteristics of the use of parking sites reserved for disabled persons and alcohol at the wheel, explain also negatively the probability of phoning while driving.

Key words: Car insurance, risky behavior, survey, trivariate probit model.

Classification JEL: C1, C31, C83, G22.

¹ ERMES (EAC 4441, CNRS), Université Panthéon Assas de Paris 2 et TEPP (FR 3126, CNRS), Fédération de recherche sur l'emploi et les politiques publiques. Adresse: Université Panthéon Assas de Paris 2, 12 Place du Panthéon, 75005 Paris, France. E-mail : meriem.maatig@u-paris2.fr

1. Introduction

L'assureur, dans sa démarche de mutualisation des risques, dispose d'informations *a priori* qui lui servent à segmenter le portefeuille en classes tarifaires. Ces classes sont supposées être homogènes du point de vue de la sinistralité (mesuré en termes de fréquence et de coût). L'hypothèse d'homogénéité des risques est une condition nécessaire pour que la politique de tarification de la compagnie d'assurance soit équitable. En effet, la présence d'informations inobservées ou inobservables (nombre de kilomètres parcourus, infractions au code de la route, état du véhicule, type de conduite,...) par l'assureur dans une classe peut venir invalider cette hypothèse d'homogénéité. Dans cette perspective, nous avons réalisé une enquête par sondage auprès de conducteurs de la région parisienne afin d'obtenir des informations sur les variables non disponibles dans le portefeuille de l'assureur, en particulier, sur les comportements à risque des conducteurs au volant.

La modélisation économétrique simple à l'aide des modèles classiques (régression logistique, modèles de comptage,...) ont un inconvénient majeur lié au biais d'estimation dû à la modélisation séparée de chaque équation: on parle de biais d'endogénéité. Afin de pallier à cet inconvénient éventuel, l'une des solutions proposées par la littérature est celle de Hausman (1978). Cette méthode consiste à élaborer un système d'équations simultanées. La caractéristique essentielle de ces modèles est que deux ou plusieurs variables endogènes sont déterminées simultanément par le modèle, comme des fonctions de variables exogènes, de variables prédéterminées, et d'aléas.

Des modèles bivariés ou multivariés ont été utilisés récemment dans de nombreux domaines. On peut citer, par exemple, pour le modèle probit bivarié, des applications dans le domaine de l'éducation, Greene (1998), dans le domaine de l'emploi, Lollivier (2001), ainsi que dans le domaine d'assurance automobile, Chiappori et Salanié (2000), Grun-Rehomme et Ben Lagha (2007). Ces derniers ont étudié un modèle bivarié où Y est la variable endogène de sinistralité et où l'autre variable endogène Z correspond aux choix du contrat de garantie. Grun-Rehomme et Ben Lagha (2007) ont prolongé l'application de leur étude empirique par le modèle ordinal bivarié. Ils ont introduit dans cette étude quatre types de formule d'assurance au lieu de deux dans l'étude de Chiappori et Salanié (2000). Ils ont montré que l'hypothèse de sélection adverse est vérifiée parmi les jeunes conducteurs qui choisissent un contrat « tous risques » variable selon le montant de la franchise.

A notre connaissance, les modèles probit bivarié et le probit trivarié ne sont pas encore utilisés sur des données françaises pour tester la relation entre les différents comportements des conducteurs au volant. L'aspect innovant de nos études est en partie lié à l'introduction d'informations non mesurables par les assureurs dans la modélisation du risque routier.

Le but de cette étude est d'avoir une spécification économétrique qui modélise simultanément les comportements à risque des conducteurs. En effet, l'existence d'un biais d'endogénéité peut nous amener à retenir, à tort, des variables explicatives dans une équation séparée. C'est pourquoi, nous avons testé l'existence de liens de causalité entre trois groupes de variables inobservées par l'assureur. Nous estimons un probit trivarié qui permet de modéliser simultanément les probabilités de l'utilisation des places de parking réservées aux handicapés, la conduite sous l'emprise de l'alcool et l'utilisation du téléphone portable au volant.

Cet article est organisé en 5 sections. Après l'introduction, la méthodologie adoptée pour mener cette recherche est exposée dans la section 2. Le modèle probit trivarié est présenté dans la section 3. Les résultats de la modélisation constituent la section 4. Cet article se termine par des conclusions (section 5) et une bibliographie.

2. Méthodologie

Les données servant de base aux résultats présentés dans ce travail sont issues d'une enquête par questionnaire que nous avons réalisé avec pour objectif d'acquérir une meilleure

connaissance sur les comportements des conducteurs au volant. L'enquête a été réalisée sur un échantillon représentatif des conducteurs de la région parisienne, de plus de 18 ans, assurés pour des véhicules 4 roues de tourisme durant l'année 2006. La passation du questionnaire a été réalisée du septembre au décembre 2006. A l'issue de cette période de recueil, 415 questionnaires exploitables ont été obtenus. Nous avons choisi d'utiliser la méthode des quotas avec un contrôle de leur constitution sur les variables sexe et âge.

Les quotas relatifs à la population de notre enquête, sur les variables « sexe » et « âge du conducteur », ont été faits à partir de données disponibles à l'INSEE (Institut National de la Statistique et des Etudes Economiques) et corrigées par les statistiques d'une mutuelle d'assurance française. Les enquêtés devaient être titulaires d'un permis de conduire et avoir conduit au cours de ces trois dernières années. Le recueil des données a été effectué très majoritairement sur la voie publique et très minoritairement par voie postale.

Ce questionnaire nous a permis de recueillir des informations sur :

- L'identification du conducteur et les caractéristiques de son véhicule : le sexe, l'âge du conducteur, la situation familiale, la profession du conducteur, l'ancienneté du permis, la formule d'assurance, la marque et le type de voiture, la puissance du véhicule et le coefficient réduction majoration ;
- Les habitudes de conduite : le port de la ceinture de sécurité, l'utilisation des places de parking réservées aux handicapés, les limitations de vitesse, les radars, les infractions au code de la route, les changements des comportements des conducteurs après les nouvelles mesures sécuritaires;
- Les attitudes à l'égard de l'utilisation du téléphone mobile au volant : nous disposons des variables relatives à cette utilisation pour répondre ou appeler avec ou sans kit mains libres tout en conduisant ;
- Les attitudes vis-à-vis de l'alcool au volant : tels que le taux d'alcoolémie maximal permis pour conduire, les comportements des conducteurs après avoir consommé de l'alcool et les attitudes face à la fatigue ;
- La sinistralité : nous disposons des déclarations du nombre de sinistres responsables et non responsables et la présence de sinistre non-déclaré durant les trois dernières années.

Le questionnaire comporte deux aspects : des informations observables par les assureurs (la sinistralité durant les trois dernières années, identification du conducteur et de son véhicule) et d'autres non observables par les assureurs (les comportements des conducteurs au volant).

3. Présentation du modèle probit trivarié

Pour étudier la relation qui pourrait exister entre les différents comportements à risque des conducteurs au volant, nous utilisons, dans un premier temps, le modèle économétrique simple, le modèle logit. Dans un second temps, nous utilisons le modèle probit trivarié, en modélisant simultanément les comportements à risque des conducteurs. Nous présentons dans cette partie le modèle probit trivarié. La première équation traite l'utilisation des places de parking réservées aux handicapés. La seconde étudie la conduite sous l'emprise de l'alcool. La dernière estime l'utilisation du téléphone portable au volant. Cette modélisation permet de tester à la fois si les comportements sont sélectifs et s'ils sont liés ou non.

Soient H , A et T les trois variables aléatoires binaires définies par :

$$H_i = \begin{cases} 1 & \text{si le conducteur } i \text{ utilise les places de parkings réservées aux handicapés} \\ 0 & \text{sinon} \end{cases}$$

$$A_i = \begin{cases} 1 & \text{si le conducteur } i \text{ conduit sous l'emprise de l'alcool} \\ 0 & \text{sinon} \end{cases}$$

$$T_i = \begin{cases} 1 & \text{si le conducteur } i \text{ utilise son téléphone au volant} \\ 0 & \text{sinon} \end{cases}$$

Avec H_i , A_i et T_i trois variables aléatoires discrètes.

On peut écrire le système suivant :

$$\begin{cases} H^* = \beta_1 X_1 + \varepsilon_1 \\ A^* = \beta_2 X_2 + \varepsilon_2 \\ T^* = \beta_3 X_3 + \gamma_1 H + \gamma_2 A + \varepsilon_3 \end{cases}$$

Où H^* , A^* et T^* sont respectivement les variables latentes continues associées aux variables aléatoires H , A et T .

Les vecteurs X_j ($j=1,2,3$) contiennent les facteurs exogènes pouvant expliquer les comportements à risque, c'est-à-dire des variables observables par l'assureur.

Les ε_i sont les termes d'erreur qui sont distribués conjointement selon une loi normale, tels que :

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix} \rightarrow N \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix} \right]$$

Les coefficients ρ_{jk} (avec $j \neq k$) reflètent les corrélations pouvant exister entre les erreurs des trois équations. Si les comportements sont indépendants, ces coefficients sont nuls. Par contre, ils sont significativement différents de zéro si les comportements sont dépendants les uns des autres.

Certaines contraintes d'identification du modèle doivent être imposées afin d'estimer tous les paramètres. La première restriction consiste à normaliser les variances à 1, soit: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2$. Du fait de la récursivité du modèle, les résidus des trois équations latentes n'étant pas indépendants, les paramètres du modèle ne peuvent pas être identifiés si les variables explicatives des différentes équations sont identiques (Maddala, 1983, pp. 117-147). Les vecteurs X_1 , X_2 et X_3 ne sont pas exactement les mêmes dans chacune des équations.

4. Résultats de la modélisation

A partir de notre enquête, nous disposons des informations non disponibles dans le portefeuille de l'assureur. Ces informations sont structurées autour de trois thèmes qui définissent autant de groupes de variables : l'utilisation du téléphone au volant, les habitudes de conduite et l'enjeu de la conduite sous l'emprise de l'alcool.

Les analyses exploratoires des données montrent que qu'une large majorité de conducteur reconnaît commettre des infractions au Code de la route. 77% des conducteurs utilisent le téléphone portable au volant avec kit mains libres, 49% utilisent le téléphone tenue à la main. Parmi les consommateurs d'alcool, 59% conduisent sous l'emprise de l'alcool. 36% ont au moins une infraction au code de la route durant les trois dernières années.

Certes, si les analyses exploratoires des données sont nécessaires et apportent quelques informations, elles doivent être complétées par une analyse multidimensionnelle pour isoler l'effet propre de chaque variable et donc avoir une idée plus approfondie sur l'intensité de la liaison entre les comportements des conducteurs au volant.

Nous avons commencé par une modélisation des effets séparés de chacune des thèmes à l'aide de modèles de régressions logistiques. Nous constatons que ces trois modèles séparés ont

permis d'identifier trois comportements à risque des conducteurs : l'utilisation des places de parking réservées aux handicapés, la conduite sous l'emprise de l'alcool et l'utilisation du téléphone portable au volant. Mais, il se peut qu'il y ait des interactions entre ces trois groupes. C'est que nous allons étudier dans la suite de cette étude.

L'analyse des résultats du modèle trivarié nous amène à deux conclusions :

- Premièrement, l'influence significative de plusieurs caractéristiques individuelles sur les comportements à risque est confirmée. Les résultats de notre modélisation montrent que seule la variable « jeune conducteur » (les conducteurs ayant une ancienneté de permis de conduire inférieure à trois ans) permet d'expliquer l'utilisation des places de parking réservées aux handicapés. Pour la conduite sous l'emprise de l'alcool, elle augmente avec les variables « profession du conducteur » et « coefficient Bonus-Malus » et diminue pour la variable « célibataire ». En ce qui concerne l'utilisation du téléphone portable au volant, elle augmente non seulement avec la variable « conduite sous l'emprise de l'alcool », mais également avec la variable « utilisation des places de parking réservées aux handicapés » et avec les infractions au Code de la route. Elle diminue pour la variable « jeune conducteur ». On constate que, les variables « utilisation des places de parking réservées aux handicapés » et « conduite sous l'emprise de l'alcool », endogènes dans les deux premières équations et exogènes dans la dernière équation, expliquent positivement l'utilisation du téléphone portable au volant.

- Deuxièmement, les coefficients de corrélation des erreurs du probit trivarié sont significativement différents de zéro entre l'utilisation des places de parking réservées aux handicapés et l'utilisation du téléphone portable au volant ainsi qu'entre la conduite sous l'emprise de l'alcool et l'utilisation du téléphone portable au volant. Il y a donc une interdépendance entre ces comportements alors que ce coefficient n'est pas significatif entre l'utilisation des places de parking réservées aux handicapés et la conduite sous l'emprise de l'alcool. Il y a aussi indépendance entre les termes d'erreurs. Les deux équations pourraient être estimées puis interprétées séparément.

Nous constatons que l'utilisation du téléphone portable au volant est expliquée non seulement par la conduite sous l'emprise de l'alcool mais aussi par l'utilisation des places de parking réservées aux handicapés. Et comme les coefficients de corrélations sont négatifs, les caractéristiques individuelles qui expliquent la conduite sous l'emprise de l'alcool et l'utilisation de ces places de parking expliquent aussi négativement la probabilité de l'utilisation du téléphone portable au volant. On met alors en évidence une corrélation entre la conduite sous l'emprise de l'alcool et l'utilisation du téléphone mobile au volant et entre l'utilisation des places de parking réservées aux handicapés et l'utilisation du téléphone.

Les probabilités prédites par le modèle probit trivarié sont utilisées pour préciser les résultats précédents.

Toutes choses égales par ailleurs, l'utilisation du téléphone portable au volant augmente avec la conduite sous l'emprise de l'alcool et augmente aussi avec l'utilisation des places de parking réservées aux handicapés et les infractions au Code de la route au cours des trois dernières années et diminue avec la variable « jeune conducteur ». Cette utilisation du téléphone augmente de 23,2 % avec la conduite sous l'emprise de l'alcool et de 19,9 % avec l'utilisation des places de parking réservées aux handicapés. Ceci pourrait être expliqué par le fait qu'un conducteur non respectueux des places de parking réservées aux handicapés (et donc des autres usagers de la route) ne respecte pas non plus les règles du Code de la route (l'utilisation du téléphone au volant). Aussi, un conducteur, s'il est sous l'emprise de l'alcool, ne peut pas contrôler ses comportements au volant.

5. Conclusion

Cette étude a révélé d'importantes informations sur le lien existant entre ces trois comportements des conducteurs. Certes les variables dont dispose l'assureur traduisent en partie les comportements des conducteurs. Mais notre étude a montré que ces comportements sont expliqués aussi par des variables inobservables ou inobservées par l'assureur.

Nous constatons que les coefficients de corrélation des erreurs (ρ) du probit trivarié sont significativement différents de zéro entre l'utilisation des places de parking réservées aux handicapés et l'utilisation du téléphone portable au volant ainsi qu'entre la conduite sous l'emprise de l'alcool et l'utilisation du téléphone portable au volant. Donc, il y a interdépendance entre ces comportements. Ces coefficients étant négatifs, les caractéristiques individuelles qui expliquent la conduite sous l'emprise de l'alcool et l'utilisation des places de parking réservées aux handicapés expliquent aussi négativement la probabilité de l'utilisation du téléphone portable au volant. Le coefficient ρ n'est significatif entre l'utilisation des places de parking réservées aux handicapés et la conduite sous l'emprise de l'alcool, les termes d'erreurs sont donc indépendants. Ces deux équations pourraient être estimées puis interprétées séparément.

Nos résultats montrent que la conduite sous l'emprise de l'alcool et l'utilisation des places de parking réservées aux handicapés augmentent la probabilité d'utiliser le téléphone portable au volant.

Les retombées opérationnelles de cette étude est de deux ordres. La première, l'identification des populations les plus réfractaires aux comportements des conducteurs au volant permettra de mieux cibler les campagnes d'informations de la prévention routière sur les risques routiers concernant les facteurs les plus significatifs qui permettent d'augmenter la probabilité d'avoir un accident. La deuxième, concerne les assureurs, cela lui permet de cibler ses démarches marketing pour faire face à la concurrence.

Plusieurs perspectives pourraient être l'objet de différentes études. La première est de tester l'endogénéité des comportements à risque des conducteurs sur la sinistralité. Pour ce faire, nous allons utiliser un modèle probit bivarié qui estime simultanément la sinistralité et les comportements à risque des conducteurs. Dans ce cas la variable sinistralité est une variable binaire (pas de sinistre, modalité de référence, ou au moins un sinistre, modalité alternative).

Une deuxième perspective est d'utiliser un modèle polytomique ordonné bivarié à équations structurelles, pour mesurer l'existence de lien entre les comportements à risque et la sinistralité, dans le cas où la variable endogène sinistralité est composée de quatre modalités.

6. Bibliographie

- [1] Chiappori P.A., Salanié B. (2000), Testing for Asymmetric Information in Insurance Markets, *Journal of Political Economy*, vol.108, n. 1, pp. 56-78.
- [2] Hausman J.A. (1978), Specification Tests in Econometrics, *Econometrica* 46(6), pp. 1251-1271.
- [3] Heckman J. J., Robb R. J. (1985), Alternative Methods for Evaluating the Impact of Interventions: An Overview, *Journal of Econometrics*, vol. 30, n°1-2, pp. 239-267.
- [4] Greene W.H. (1998), Gender economics courses in liberal arts colleges: further results, *Journal of Economic Education*, vol. 29, pp. 291-300.
- [5] Grun-Réhomme M., BenLagha N. (2007), Choix de contrat et sinistralité chez les jeunes conducteurs, *Revue Assurances et gestion des risques*, vol.74-4, 505-532, Montréal, Canada.
- [6] Lollivier S. (2001), Endogénéité d'une variable explicative dichotomique dans le cadre d'un modèle bivarié, *Annales d'économie et de statistique*, vol. 62, pp. 251-269.
- [7] Maddala G. (1983), Limited Dependent and Qualitative Variables in Econometrics, *Cambridge University Press*.