The Informational Content of Household Decisions with Applications to Insurance Under Adverse Selection

by Georges Dionne,
Christian Gouriéroux and Charles Vanasse

Working Paper 98-02
January 1998

ISSN : 1206-3304

Comments by P.A. Chiappori, A. Snow, and participants to the meetings of the 24th Seminar of the European Group of Risk and Insurance Economists (Paris), the Théma-Delta Seminar on insurance in Paris, and the Société canadienne de science économique (Montréal) are acknowledged.
This research was financed by the FFSA, CRSH Canada, CREST, FCAR Quebec and the Risk Management Chair at HEC Montreal.

The Informational Content of Household Decisions with Applications to Insurance Under Adverse Selection

Georges Dionne, Christian Gouriéroux and Charles Vanasse

Georges Dionne holds the Risk Management Chair and is professor of finance at École des HEC.

Christian Gouriéroux is Director, Laboratory Finance–Assurance CREST and Research Associate, CEPREMAP.

Charles Vanasse is research professional, CRT, Université de Montréal.

Abstract

We discuss how to detect the informational content of household decisions among the explanatory variables of econometric models. Two applications to the choice of automobile insurance contracts and the demand for life insurance are provided. We show that the information provided by additional decision variables is rather weak and often non significant. In particular, there is no residual adverse selection when appropriate risk classification is applied in automobile insurance; so, the choice of a deductible does not reveal information about individual risk. Similarly, the choice of a particular portfolio does not add information on risk aversion in life insurance contracting.

Keywords : *Informational content, household decisions, automobile insurance, demand for life insurance, residual adverse selection, risk classification, deductible, risk aversion, conditional independence and endogenous choice.*

JEL number: C25, D81, G11, G22.

Résumé

Nous proposons des procédures pour tester l'utilité de prendre en compte les décisions des ménages parmi les variables explicatives des modèles économétriques. Deux applications de la démarche au choix d'assurance automobile et à la demande d'assurance des ménages sont fournies. Nous montrons que l'information obtenue par l'introduction de variables de décision additionnelles est non importante et souvent non significative. En particulier, il ne reste pas d'antisélection résiduelle lorsqu'une classification des risques appropriée est utilisée; par conséquent, le choix de la franchise à l'intérieur des classes de risque ne révèle pas d'information sur les risques individuels. De la même façon, le choix d'un portefeuille particulier n'ajoute pas d'information sur la riscophobie dans le choix de contrat d'assurance vie.

Mots clés : *Décisions des ménages, assurance automobile, demande d'assurance vie, information, antisélection résiduelle, classification des risques, franchise, riscophobie, indépendance conditionnelle et choix endogène.*

Codes JEL : C25, D81, G11, G22.

1. Introduction

Under asymmetrical information, the empirical studies on household behaviour concerning financial products or insurance contracts are generally concerned by the prediction of some individual endogenous variable related to individual risk or demand. Then the prediction formula is used to classify (score) the individuals and to construct homogenous subpopulations.

The variable of interest is often predicted by means of a nonlinear regression model if the choice is qualitative, including as explanatory variables some exogenous characteristics such as age, occupation, housing location, income level... But other variables summarizing endogenous choices of the agents may also be introduced and an important question concerns the additional information they provide.

For instance, the type of selected automobile insurance contract, i.e. the level of deductible, can be introduced to predict the number and the cost of car accidents of the insured. The choice of a graduated monthly payment instead of a constant monthly payment or the choice of a colateral can provide information on the future rate of no payment. The type of held financial assets in the individual portfolio may improve the prediction of the holding of life insurance since they can approximate risk aversion.

The theoretical arguments proposed for the introduction of such decision variables among the regressors are twofold. First the individual may possess more information than the econometrician or the insurer on his risk (or risk aversion), and part of this additional information may be revealed through some decision variables. This is the standard argument of adverse selection, where the choice of an automobile insurance contract with a large deductible reveals a better risk. [Rothschild and Stiglitz (1976) and Wilson (1977). See Dionne and Doherty (1992) for a survey].

Secondly the individual may take joint decisions, and in such a case the partial analysis of one kind of decision irrespective of the other ones may be inefficient. The joint decision of both life or liability insurance and securities is a good example since the choice of a particular portfolio may reveal information about risk aversion[1].

Of course these two arguments may be mixed. Moreover, in the case of moral hazard, an additional individual specific information, the individual's e®ort, can be simultaneously chosen along with other assets or insurance contracts. This dimension of the problem will not be discussed explicitely in this article [see however Dionne, Gourieroux and Vanasse (1998) and Chassagnon and Chiappori (1996)].

In Section 2 we discuss the notion of conditional independence and explain how it can be used in our framework. We de¯ne some measure of the informational content of these decision variables, we introduce test statistics of the null hypothesis of no informational content, and we study how these notions and statistics depend on the initial exogenous information.

This conditional dependence analysis is usually performed in practice in a parametric framework, where the model is a priori constrained. This practice may induce spurious conclusions, since it is di±cult to distinguish between an informational content of the decision variables and an omitted nonlinear e®ect of the initial exogenous variables. We discuss in Section 3 a pragmatic way for avoiding this di±culty, which consists of introducing jointly among the regressors the decision variables and their expected values computed from the initial information.

In Section 4, this approach is applied to the analysis of automobile accidents in Quebec and to the prediction of the demand for life insurance in France. The lesson from these examples is that the additional information provided by the decision variables is rather weak and often non signi¯cant as soon as the nonlinear e®ect of the initial exogenous variables have been introduced in a suitable way. Other conclusions are summarized in Section 5.

2. Conditional dependence and independence

The problem of additional information may be treated by means of conditional dependence. In this section, we recall the main results on this notion [see e.g. Gourieroux-Monfort (1995) Volume 2 p 458-475]. We denote by Y the endogenous variable of interest, by X the K initial exogenous variables and by Z the L decision variables.

2.1 Conditional independence

The decision variables provide no additional information if and only if the prediction of $g(Y)$ based on X and Z jointly, coincides with its prediction based on X alone. In a nonlinear framework this condition has to be valid for any transformation g of the Y variable and may be written in terms of conditional probability :

$$l(Y|X,Z) = l(Y|X), \tag{1}$$

where $l(.|.,.)$ denotes a conditional pdf.

The previous condition admits equivalent forms :

$$l(Z|X,Y) = l(Z|X), \tag{2}$$

and,

$$l(Y,Z|X) = l(Y|X)l(Z|X). \tag{3}$$

From (3), we deduce the symmetry in Y and Z of the conditional independence, and, from (2), we see that this is equivalent to the absence of additional informational content of the Y variable for predicting the decision variables.

2.2 Measure of conditional dependence

It is also standard to define valid measures of conditional dependence in a nonlinear framework. These measures are based on the so-called information criterion, first evaluated conditionally to X, and then possibly averaged on the values of the exogenous variables. More precisely, we define :

$$\begin{aligned} M(Z,Y|X) &= E\left[\log \frac{l(Y|X,Z)}{l(Y|X)} \Big| X\right] \\ &= \int\int \log \frac{l(y|X,z)}{l(y|X)} l(y,z|X)\,dy\,dz. \end{aligned} \tag{4}$$

It is known that :

$$\begin{aligned} M(Z,Y|X) &= -E\left[E\left(\log \frac{l(Y|X)}{l(Y|X,Z)} \Big| X,Z\right) \Big| X\right] \\ &\geq -E\left\{\log E\left(\frac{l(Y|X)}{l(Y|X,Z)} \Big| X,Z\right) \Big| X\right\} \quad \text{(from the convexity inequality)} \\ &= 0. \end{aligned}$$

Moreover this non negative measure vanishes if and only if $l(Y/X;Z) = \lambda(X)l(Y/X)$; for some function λ. Since the pdf has unit mass, this condition is equivalent to : $l(Y/X;Z) = l(Y/X)$, i.e. to conditional independence.

$M(Z;Y/X)$ is a dependence measure between Z and Y, computed for the different homogenous groups of individuals defined from the exogenous variables.

These measures may be summarized by a more global one corresponding to the whole population of interest, by averaging on X :

$$\begin{aligned} \bar{M}(Z;Y/X) &= E \log \frac{l(Y/X;Z)}{l(Y/X)} \\ &= E\left[E \log \frac{l(Y/X;Z)}{l(Y/X)}\right] \\ &= E_X M(Z;Y/X). \end{aligned}$$

2.3 The effect of exogenous information

The value of introducing the additional decision variables is contingent to the initial exogenous information. A question of interest is : What happens if for instance this information is increased ?

Let us distinguish two sets of exogenous variables $X = (X_0; X_1)$. We get :

$$\frac{l(Y/X;Z)}{l(Y/X)} = \frac{l(Y/X_0;Z)}{l(Y/X_0)} \frac{l(Y/X_0;X_1;Z)}{l(Y/X_0;Z)} \frac{l(Y/X_0)}{l(Y/X_0;X_1)}.$$

By taking the logarithm and the expectation of both sides, we derive a decomposition formula of the conditional dependence measure :

$$\bar{M}(Z;Y/X) = \bar{M}(Z;Y/X_0) + \bar{M}(X_1;Y/X_0;Z) - \bar{M}(X_1;Y/X_0); \qquad (5)$$

where the terms $\bar{M}$ are nonnegative.

The additional information contained in the decision variables may increase or decrease depending on the new variables X_1 introduced in the exogenous information. In particular we may likely select different exogenous information sets, more or less informative, and such that the conditional independence hypothesis is satisfied.

3. Conditional dependence or misspecified structure

3.1 Null and alternative hypotheses

The conditional independence hypothesis can be tested by either nonparametric or parametric techniques. This latter approach is generally retained for applications to finance and insurance decisions. It requires a preliminary parametric modelling for the conditional distribution of the endogenous variable of interest Y given the different explanatory variables X and Z. To simplify the presentation we consider the case of dichotomous variables[2] Y and $Z_l, l = 1, \dots, L$. Typically a parametric formulation gives the conditional probability :

$$P[Y = 1/X, Z] = F(g(X, b) + c'Z), \tag{6}$$

where F and g are given functions; F is a cumulative distribution function, and b and c are unknown parameters.

In this framework the conditional independence between Y and Z given X is characterized by the constraint $c = 0$.

Under this null hypothesis $H_o = \{c = 0\}$, we get :

$$P[Y = 1/X, Z] = P[Y = 1/X] = F[g(X, b_o)],$$

where b_o is the true value of the parameter.

The null hypothesis may be rejected as a consequence of either conditional dependence :

$$P[Y = 1/X, Z] \neq P[Y = 1/X],$$

or a misspecified structural form :

$$P[Y = 1/X] \neq F[g(X, b)], \forall b.$$

This second reason may be avoided by selecting a sufficiently smooth specification, including cross effects. This is the point we are now going to discuss.

3.2 Example of linear scoring function

In practice the scoring function $S(X;Z) = g(X;b) + c'Z$ if often written as a linear function of the explanatory variables $S(X;Z) = b'X + c'Z$, i.e. without introducing cross effects of the individual characteristics. Jointly some similar specifications may be introduced for the $Z_l, l = 1, \dots, L$ variables :

$$P[Z_l = 1/X] = F(a_l'X).$$

Moreover we assume that the Z_l variables are independent.

Let us now consider this modelling when the conditional dependence is small : $c \simeq 0$. The conditional distribution of Y given only the exogenous variables X is :

$$\begin{aligned}
& P[Y = 1/X] \\
= \; & \sum_{z_1=0}^{1} \dots \sum_{z_L=0}^{1} \left\{ \prod_{l=1}^{L} \left(F(a_l'X)^{z_l}(1 - F(a_l'X))^{1-z_l}\right) F\Big(b'X + \sum_{l=1}^{L} c_l z_l\Big) \right\} \\
\simeq \; & F(b'X) + \dot{F}(b'X) \sum_{z_1=0}^{1} \dots \sum_{z_L=0}^{1} \prod_{l=1}^{L} \left(F(a_l'X)^{z_l}(1 - F(a_l'X))^{1-z_l}\right) \sum_{l=1}^{L} c_l z_l \\
= \; & F(b'X) + \dot{F}(b'X) \sum_{l=1}^{L} c_l F(a_l'X) \\
\simeq \; & F\Big(b'X + \sum_{l=1}^{L} c_l F(a_l'X)\Big),
\end{aligned}$$

where $\dot{F}$ is the derivative of F.

The general form of the conditional distribution $P[Y = 1/X]$ is very different from the linear scoring corresponding to the null hypothesis[3]. The linear introduction of the decision variables $Z_l, l = 1, \dots, L$ inside the scoring function is an artificial way of introducing cross effects of the X variables, through the expectations $F(a_l'X), l = 1, \dots, L$. Indeed the second order derivative of the score with respect to variables X_1, X_2 (say) is equal to : $\frac{\partial^2 (b'X + \sum_{l=1}^{L} c_l F(a_l'X))}{\partial X_1 \partial X_2} = \sum_{l=1}^{L} c_l \, a_{1l} \, a_{2l} F''(a_l'X)$, and is generally different from zero. This example shows that the linear scoring functions are too constrained and that the rejection of the null hypothesis $\{c_l = 0\}, \forall l$, will likely detect the omission of cross-effects.

3.3 How to smooth the linear scoring functions ?

The modelling with linear scoring functions can be easily extented to avoid the main part of the previous difficulty. We simply have to consider a modified specification :

$$
\begin{aligned}
& P[Y = 1/X, Z] \\
= \quad & F[b'X + \textstyle\sum_{l=1}^{L} d_l F(a_l' X) + \sum_{l=1}^{L} c_l Z_l],
\end{aligned}
$$

in which the decision variables are introduced jointly with their expectations conditional to X. The introduction of predictions of endogenous variables inside the explanatory variables is similar to the idea followed for defining Regression Specification Error Test [RESET] [Ramsey (1969), Godfrey (1988) p 106]. The difference is that in our case the introduced prediction concerns other endogenous variables.

4. Applications

We will apply the previous approach by comparing models in which the additional variables introduced in the linear scoring are the $Z_l, l = 1, \ldots, L$ only, to models containing both these variables and their expectations. We will see that spurious conditional dependence may be exhibited if we omit the expectations [see Puelz-Snow (1994), for such interpretation and Chiappori-Salanié (1996) for a different approach to that proposed in this article].

4.1 Joint analysis of automobile accidents distribution and deductible choice

This type of analysis has been performed by Puelz-Snow (1994). The authors considered an ordered logit formulation for the deductible choice (Z) in which the observed number of accidents (Y) was introduced among the explanatory variables. The estimated coefficient of the Y variable is significant and they concluded to the presence of adverse selection (i.e. of conditional dependence between Y and Z). It can be noted that the test procedure has been based on the indirect characterization (2) and not on the direct one (1). Such a practice may be interpreted as the description of what will be the decision of the individual if he had perfect knowledge of the future risk.

We will show that the derived conclusion is likely a spurious e®ect, due to the too constrained form of the exogenous e®ects. In fact, the linear speci¯-cation of their ordered logit model contained only few variables. For this purpose, we consider the indirect form of the conditional distribution of Z given Y and X, in which we introduce linear e®ect of the X variables plus nonlinear e®ect through an expected value of the number of accidents. This expectation is based on a preliminary negative binomial model estimated with only the X as explanatory variables [Gourieroux-Monfort-Trognon (1984), Dionne-Vanasse (1992), Lemaire (1995), Dionne et al. (1997), Pinquet (1997)], [see Appendix 1 for the estimated model and Dionne-Gourieroux-Vanasse (1996) for more details].

The data come from a large private insurer in Quebec. Di®erent contracts corresponding to various levels for a straight deductible are proposed, but the deductible choice does matter for only two levels of deductible $250 and $500, and the choice of $500 was done only by about 4% of the overall portfolio.

Figure 4.1 and Table 4.1 indicate that the proportion of individuals who choose the $500 deductible varies between risk classes. These risk classes are not directly observable and were built up from observable variables such as age, sex, territory... The question of interest is the following: do these choices of deductibles reveal private information on individual risk? To answer, we did the following analysis for the classes 4 to 19, where the $500 deductible choice is signi¯cant.

(Figure 4.1 and Table 4.1 about here.)

The main exogenous variables introduced in the econometric speci¯cations of the deductible choice equation (Z) are : Age of the principal driver ; SexF (1 if the principal driver is a female) ; Gj a group of 8 dummy variables representing car classi¯cation groups of the insurer ; Occasional young male (YMALE) driver, if there is such a driver in the household. All these variables and others have been introduced since they are used in the tari¯cation of the insurance company. Moreover, as in Puelz and Snow (1994), the number of current accidents N (acc) is introduced in the ¯rst model while, in the second model, the expected number of accidents E (acc) is added. We did also control for risk aversion by introducing wealth proxy variables Wi that indicate the chosen liability insurance coverage. Finally, a price variable (GD) for the $500 deductible was obtained from the tari¯cation book of the insurer : This is the rebate for the passage from the $250 to the $500 deductible. [see Appendix 1 for the whole list of variables].

In a ¯rst step, probit models for the choice of a deductible of $500 have been estimated for all drivers, ¯rst with the number of claims (over $500) only (Model 1), and then jointly with the expected number of accidents (Model 2)[4]. The speci¯cations of the two models do not contain all the available classi¯cation variables as in Puelz and Snow (1994). More variables will be considered in Model 3. The ¯rst columns of Table 4.2 give the estimated coe±cients and the second ones the associated student statistics.

(Table 4.2 about here.)

The results indicate clearly that when the model is not correctly speci¯ed a false conclusion can be made about the presence of residual adverse selection in automobile insurance. Model 1 suggests that Y and Z are correlated or that the null hypothesis of conditional independence is rejected implying the presence of residual adverse selection in the risk classes. Indeed, as in Puelz and Snow (1994), we obtain that the coe±cient of N (acc) is negative and signi¯cant, indicating that those who experience more accidents choose the low deductible. This conclusion is, in fact, not appropriate. When we add the expected number of accidents (E (acc)) in the model, the coe±cient of N (acc) is no more signi¯cant[5]. This means that when we take into account of the nonlinearity of the risk classi¯cation variables through E (acc), the residual adverse selection in the risk classes vanishes. In other words, by an appropriate risk classi¯cation procedure, the insurer, when using observable variables, is able to control for adverse selection and does not need any additional self-selection mechanism. [See Crocker and Snow (1986) for a theoretical analysis of risk classi¯cation under adverse selection and Dionne and Doherty (1992) for a review of the di®erent insurance contracting models]. Finally, Model 3 in Table 4.3 shows that we can eliminate the E (acc) variable by using more classi¯cation variables as insurers do[6]. Even the proxies for wealth variables (Wi), used to control for risk aversion, are no more signi¯cant while two categories were signi¯cant in Models 1 and 2.

(Table 4.3 about here.)

4.2 Holding of life insurance in France

The second application concerns the portfolio allocation by French households. It is well known that individual portfolios are not well diversi¯ed

[Michael-Hamburger (1968), Shorrocks (1982), King-Leape (1984), Gourieroux-Tiomo-Trognon (1996)]. This result is contrary to the standard ¯nancial theory [Markowitz (1992)], but can be explained by transaction costs, the impossibility to have short positions, the illiquidity of a number of assets such as housing, human capital, the commercial e®orts of the banks and insurance companies and by asymmetrical information in some markets such as life insurance. Therefore it is useful to begin a study of portfolio allocation by considering qualitative features such as the type of assets introduced in the portfolio.

In the traditional literature on life insurance and adverse selection [see Villeneuve (1996) for a recent literature review], it is shown that risk classi¯cation variables are very useful to approximate the individual risks. However, when individuals di®er also in their risk aversion more instruments are necessary to make prediction of insurance demand. For example, interaction variables with income and total wealth (when available) can be used to increase the number of risk classes. Here we will show that the decision variables of other ¯nancial securities do not provide strong additional information when the traditional exogenous variables are introduced in an appropriate way. In other words, residual risk aversion can be captured by appropriate classes of insureds.[7]

The data corresponds to a sample of French households observed for the year 1995. Di®erent informations are available on individual characteristics, and on the type and amount of assets they have in their portfolio. These assets have been grouped in four classes, i.e. liquid assets [Bank account, short term T-bond, short term mutual fund], home buyer saving scheme (HBS), stocks and bonds, and life insurance. The ¯scal conditions for life insurance in France explain its return and why it is a competitor to more traditional assets. In Table 4.4 we give some information on the diversi¯cation level of the studied portfolios.

(Table 4.4 about here.)

We are interested in the prediction of life insurance demand. Under asymmetrical information, this demand is function of the non-observable individual risks (approximated by exogenous risk classi¯cation variables), their risk aversion and their demand for other assets. In this study, the other decision variables concern the holding of three other categories of assets. The exogenous variables for risk classi¯cation are age and $(\text{age})^2$ of the head of household to account for life cycle e®ect, current income, total ¯nancial wealth, sex (1 for woman); occupation : superior, intermediate, employees, workers, retired, unactive (others) ; type of district : rural, between 2000 and 20 000 inhabitants, between 20 000 and 100 000, more than 100 000, (Paris); education level : (primary), technical, Baccalaureat, graduate and post graduate; position for housing : owner, lender, (free disposal); type of household : (alone), one adult and children, couple with two active people without child, couple with two active people with children, couple without activity, couple with one active people. This set of variables is used ¯rstly to estimate separately logit models for the three di®erent decision variables, then they are reintroduced in the logit formulation for the holding of life insurance. The two estimated logit regressions for life insurance with the decision variables only and jointly with their expectations are given in Table 4.5. For each model the ¯rst column gives the estimated coe±cients and the second one the corresponding Wald chi-square statistics, whose critical value is about 6.3 at 99%. All the other regressions for the other decision variable are available upon request.

(Table 4.5 about here.)

As in the previous example, without introducing the expected decision variables, all the choice variables (Liquid asset, HBS and Stock and Bond) are highly signi¯cant. But they become almost non signi¯cant when their expectations are introduced. From the analysis of the ¯rst logit model, (Model 4) we may get the impression of some dependence between the choices conditional to the exogenous variables, whereas this is mainly due to the omission of some cross-e®ects taken into account by the expected variables of the second logit speci¯cation (Model 5). The substitution e®ects are conditional to the initial information. The coe±cients of the expected variables indicate that the more risk averse decision makers (who hold liquid asset and HBS) have a higher life insurance demand than the less risk averse (who hold stock and bond). But, as in the previous example, since these coe±cients were obtained from observable variables, the result also means that there is no signi¯cant residual risk aversion in the portfolio. Finally, as in the previous example, on can show that, by appropriate use of other classi¯cation variables or by interactions of the available ones, the expected variables will become themselves no more signi¯cant.

5. Conclusion

In this paper, we have introduced the notion of conditional independence and showed how it can be applied to our framework of individual choices under asymmetrical information. We have shown that spurious conclusions can be drawn in di®erent applications since it is di±cult to separate the information content of a decision from omitted nonlinear e®ects of initial exogenous random variables.

Two applications to insurance decisions under adverse selection were presented. In the ¯rst one, we analyzed jointly the automobile accidents distribution and the deductible choice. One prediction in the literature is that high risk individuals should choose small deductibles inside risk classes when there remains asymmetrical information. We showed, however, that risk classi¯cation is su±cient in the sence that there is no residual adverse selection on risk types in the automobile insurance portfolio studied. We obtained a similar conclusion for the variables used to measure risk aversion.

In the second example, we considered the joint decision of holding life insurance and other ¯nancial assets. In this example, since we do not have information on individuals risks, the asymmetrical information of interest is risk aversion. The decision on other assets may reveal information on risk aversion. Those who hold positions in more risky assets should be less risk averse and hold less life insurance. But assets decisions variables are almost not signi¯cant when their expectations on observable variables are introduced. There is again no strong residual adverse selection on risk aversion types in the life insurance portfolio considered.

Of course, there is (marginal) adverse selection in these markets. The message of this article is that appropriate combinations of exogenous variables are su±cient to capture the asymmetric information. In other words, when appropriate observable characteristics are used, no other self-selection mechanism seems necessary. However, the expected values of the decision variables (or di®erent cross combinations of the observables) should be used to take into account of non-linearities.

REFERENCES

Chassagnon, A., and P. A. Chiappori (1996) : "Insurance Under Moral Hazard and Adverse Selection : the Case of Pure Competition", Working Paper 28, DELTA.

Chiappori, P., and B., Salani¶e (1996)a : "Asymmetric Information in Automobile Insurance Markets : An Empirical Investigation", Mimeo , DELTA.

Chiappori, P., and B., Salani¶e (1996)b : "Empirical Contract Theory : The Case of Insurance Data", CREST. European Economic Review (forthcoming).

Crocker, K.J., and A. Snow (1986) : "The E±ciency E®ects of Categorical Discrimination in the Insurance Industry", Journal of Political Economy, 94, 321-344.

Dionne, G. and N. Doherty (1992) :"Adverse Selection in Insurance Markets : A Selective Survey", in G. Dionne (ed), Contributions to Insurance Economics, Boston, Kluwer Academic Press.

Dionne, G., Gagn¶e, R., Gagnon, F. and C. Vanasse (1997), "Debt, Moral Hazard and Airline Safety: An Empirical Evidence", Journal of Econometrics, 79, 379-402.

Dionne, G., Gouri¶eroux C., and C., Vanasse (1998) :"Evidence of Adverse Selection in Automobile Insurance Markets". Forthcoming in Dionne G. and C.L. Nadeau (eds.), "Automobile Insurance", Kluwer Academic Press.

Dionne, G., and C., Vanasse (1992) : "Automobile Insurance Ratemaking in the Presence of Asymmetrical Information", Journal of Applied Econometrics, 7, 149-165.

Godfrey, L.G. (1988) : "Misspeci¯cation Tests in Econometrics; the Lagrange Multiplier Principle and Other Approach", Cambridge University Press, 252 p.

Gouri¶eroux, C. (1997), "The Econometrics of Insurance", Mimeo, CREST and CEPREMAP. Geneva Papers on Risk and Insurance Theory (forthcoming).

Gouri¶eroux, C., and A., Monfort (1995) : "Statistics and Econometric Models", Cambridge University Press, volume 2, 458-475.

Gouriéroux, C., Monfort, A., and A., Trognon (1984) : "Pseudo Maximum Likelihood Methods : Application to Poisson Models", Econometrica, 52, 701-721.

Gouriéroux, C., Tiomo, A., and A., Trognon (1996) : "The Portfolio Composition of Households : Some Evidence from French Data", CREST.

Hamburger, M.J., (1968) : "Household Demand for Financial Assets", Econometrica, 56, 97-118.

King, M., and J., Leape (1984) : "Wealth and Portfolio Composition : Theory and Evidence", WP. NBER 2468.

Lemaire, J. (1995) :"Bonus-Malus Systems in Automobile Insurance", Kluwer, Boston.

Markowitz, H. (1992) : "Portfolio Selection : Efficient Diversification of Investment", 2nd ed., Wiley, New-York.

Mayers, D., and C.W. Smith (1983), "The Interdependence of Individual Portfolio Decisions and the Demand for Insurance", Journal of Political Economy, 91, 304-311.

Mc Fadden, D. (1973) :"Conditional Logit Analysis of Qualitative Choice Behavior", in P. Zarembka (ed), Frontiers in Econometrics, New-York, Academics Press.

Murphy, K.M. and R.H. Topel (1985), "Estimation and Inference in Two-Step Econometric Models", Journal of Business & Economic Statistics, 3, 370-379.

Pagan, A. (1984) : "Econometric Issues in the Analysis of Regressions with Generated Regressors", International Economic Review, 25, 221-247.

Pinquet, J. (1997), "Experience Rating for Heterogeneous Models", Working paper 9725 THEMA, 48 pages. Forthcoming in Handbook of Insurance, G. Dionne (ed), Kluwer Academic Press.

Puelz, R., and A., Snow (1994) : "Evidence on Adverse Selection : Equilibrium Signalling and Cross Subsidization in the Insurance Market", Journal of Political Economy, 102, 236-257.

Ramsey, J. (1969) : "Test for Specification Errors in Classical Linear Least-Squares Regression Analysis", Journal of the Royal Statistical Society, B, 31, 350-371.

Rothschild, M., and J. Stiglitz (1976) : "Equilibrium in Competitive Insurance Markets : An Essay on the Economics of Imperfect Information", Quarterly Journal of Economics, 90, 629-649.

Shorrocks, A. (1982) : "The Composition of Asset Holdings in the United Kingdom", Economic Journal, 92, 268-284.

Villeneuve, B. (1996) : "Essais en économie de l'assurance" Ph.D. thesis, EHESS, 269 pages.

Wilson, C. A. (1977) : "A Model of Insurance Markets with Incomplete Information", Journal of Economic Theory, 16, 167-207.

Notes :

[1] On the joint demand of liability insurance and portfolio assets see, for example, Mayers and Smith (1983).On insurance decision in presence of adverse selection with di®erent risk averse individuals, see Villeneuve (1996).

[2] The presentation can be extended to the case of discrete variables. In fact, in one application, Y is a count variable.

[3] The previous expansion shows that the conditional distribution of Y and X may be derived simply by instrumenting the endogenous decision variables inside the scoring function. This result is only valid locally (i.e. for ' 0), and such a practice will lead in general to a misspeci¯ed formulation for P[Y = 1=X] and to unconsistent estimators of the c parameters [see Pagan (1984)]

[4]As in Puelz and Snow (1994), we did not consider the claims between $250 and $500 since they are not observable for those who choose the higher deductible.

[5] Our second-step regression (deductible choice) contains a stochastic regressor, E(acc). It is well known that such a two-step procedure yield consistent estimates of the coe±cients. However, the second-step estimated standard errors based on this procedure will generally be biased. Murphy and Topel (1985) proposed a general correction to the estimated variance matrix in order to correct standard errors in two-stage estimation. The application of the proposed correction (Murphy and Topel, p.377) did not change our results: signi¯cant (non-signi¯cant) coe±cients remain the same. These supplementary results are available upon request from the authors.

[6] We did also estimate a model with N(acc) only and more classi¯cation variables than in Model 1. Again, N(acc) became not signi¯cant. Results are available from the authors.

[7] Here the residual adverse selection on risk types cannot be studied since we do not have access to the data on accidents.

Appendix 1
Definition of variables for automobile insurance example

AGE : Age of the principal driver

SEXF : Dummy variable equal to 1, if the principal driver is a female.

MARRIED : Dummy variable equal to 1, if the principal driver of the car is married.

Z : Dummy variable equal to 1, if the deductible is $ 500 [equal to 0 for a $ 250 deductible].

T1 to T22 : Group of 22 dummy variables for territories. The reference territory T1 is the center of the Montreal island.

G8 to G15 : Group of 8 dummy variables representing the tariff group of the used car. The higher the actual market value of the car, the higher the group. G8 is the reference group.

CL4 to CL19 : Driver's classes, according to age, sex, marital status, use of the car and annual mileage. The reference class is 4. (See Figure 4.1 for their identifications.)

NEW : Dummy variable equal to 1 for insured entering the insurer's portfolio.

YMALE : Dummy variable equal to 1, if there is a declared occasional young male driver in the household.

AGEAUTO : Age of the car in years.

N (acc) : Observed number of claims [for accidents where the loss is greater than $500] (range 0 to 3).

E (acc) : Expected number of accidents obtained from the negative binomial regression estimates.

GD : Marginal price (rebate) for the passage from the $250 to the $500 deductible. This amount is negative and comes from the tariff book of the insurer.

W1 to W5 : Chosen limit of liability insurance. W1 is the reference limit.

Alpha : Overdispersion parameter of the negative binomial distribution.

Table A1
Negative Binomial on Automobile Accidents

Variable	Coefficient	T-ratio
Intercept	–1.86280	–6.832
SEXF	–0.27216	–2.294
MARRIED	0.11436	0.959
AGE	–4.47E–03	–0.763
NEW	0.31644	2.871
Group of vehicles		
G9	–4.58E–02	–0.381
G10	–1.78E–03	–0.011
G11	0.12375	0.447
G12	0.27727	0.833
G13	0.60915	1.708
G14	–7.47E–02	–0.112
G15	6.26E-02	0.078
Territory		
T2	–0.36545	–0.748
T3	–0.28546	–0.973
T4	–0.75719	–2.406
T5	–6.77E–02	–0.279
T6	–0.51594	–1.412
T7	–0.37108	–1.787
T8	–0.94753	–1.888
T9	–0.19458	–0.632
T10	1.32E–02	0.033
T11	–0.76729	–2.989
T12	–0.72699	–1.431
T13	–0.18672	–0.551
T14	–0.57162	–2.386
T15	0.22855	0.552
T16	–0.95952	–1.430
T17	0.47768	0.861
T18	–0.63773	–1.776
T19	–0.96049	–3.068
T20	–0.96003	–2.694
T21	–0.44106	–1.641
T22	–0.47611	–1.916
Alpha	0.36905	1.299
Number of observations	4772	
Log-Likelihood	–1515.045	

Table 4.1
Deductibles and Risk Classes

Class	$250 deductible		$500 deductible	
	N	% of class	N	% of class
1	14,015	96.32%	535	3.68%
2	13,509	96.53%	486	3.47%
3	4,538	96.49%	165	3.51%
4	756	81.82%	168	18.18%
7	1,515	92.66%	120	7.34%
8	11	68.75%	5	31.25%
9	287	83.19%	58	16.81%
10	5	100.00%	0	0.00%
11	53	57.61%	39	42.39%
12	164	69.79%	71	30.21%
13	308	74.94%	103	25.06%
18	175	87.94%	24	12.06%
19	855	93.96%	55	6.04%
Total	36,191	95.19%	1,829	4.81%

Figure 4.1

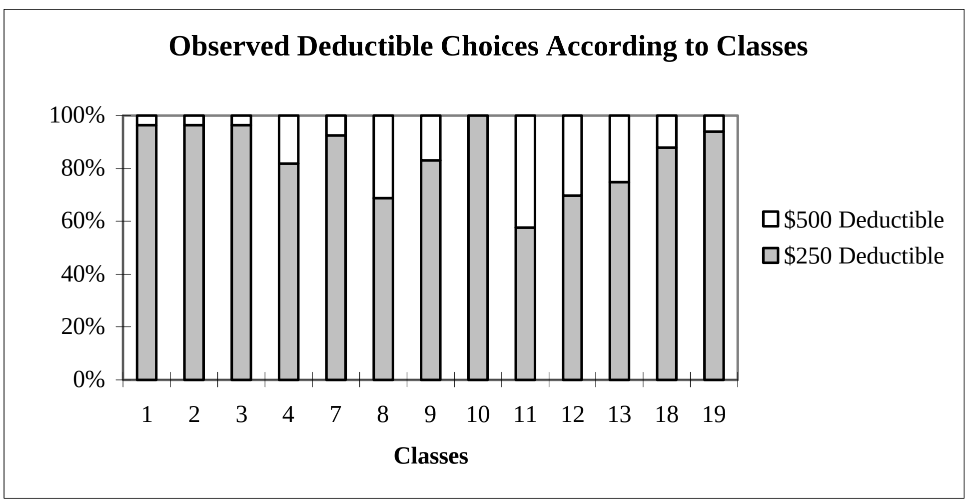

Table 4.2
Probit on Deductible Choice
(1 if 500$ deductible)

Variable	**Model 1** Conditional on the number of claims		**Model 2** Conditional on the number of claims and expected number of claims	
	Coefficient	**T-ratio**	**Coefficient**	**T-ratio**
Intercept	–0.7505	–5.006	–0.4884	–3.111
Acc	–0.1579	–1.983	–0.1151	–1.436
E(acc)			–5.4637	–6.524
GD	–0.0099	–5.275	–0.0150	–7.299
SEXF	–0.5097	–8.296	–0.5968	–9.426
AGE	–0.0251	–7.975	–0.0241	–7.681
Liability limit				
W2	–0.0133	–0.177	–0.0360	–0.474
W3	–0.2016	–1.872	–0.2016	–1.860
W4	0.0115	0.172	0.0427	0.635
W5	–0.2337	–2.990	–0.1634	–2.063
Group of vehicles				
G9	0.1484	2.683	0.1266	2.268
G10	0.2428	3.359	0.2475	3.410
G11	0.4242	3.267	0.4905	3.754
G12	0.6934	4.346	0.8398	5.165
G13	0.7974	4.485	1.3053	6.709
G14	1.1424	4.937	1.0745	4.675
G15	1.0582	3.541	1.0690	3.551
YMALE	0.1127	0.734	0.0589	0.384
Number of observations	4,772		4,772	
Log-likelihood	–1,735.406		–1,713.091	

Table 4.3
Probit Estimates on Deductible Choice

Variable	Model 3 Conditional on the number of claims, expected number of claims and additional risk classification variables	
	Coefficient	T-ratio
Intercept	–0.47151	–0.777
Acc	–0.11166	–1.352
E(acc)	–2.62320	–0.772
GD	–0.00195	–0.530
SEXF	–0.08582	–0.571
AGE	–0.01352	–2.694
Liability limit		
W2	0.06720	0.837
W3	–0.12067	–1.054
W4	0.11830	1.621
W5	–0.03462	–0.395
Group of vehicles		
G9	0.16806	2.799
G10	0.29861	3.928
G11	0.48917	3.445
G12	0.75350	3.885
G13	1.07560	3.126
G14	1.10850	4.673
G15	1.29840	4.211
YMALE	0.29254	1.795
Territory		
T2	–0.12335	–0.357
T3	0.15908	0.775
T4	–0.01370	–0.042
T5	–0.18685	–1.202
T6	–0.32644	–1.100
T7	–0.55344	–2.595
T8	–0.21743	–0.577
T9	–0.85540	–3.372
T10	–0.38619	–1.391
T11	–0.14505	–0.466
T12	–0.20954	–0.607
T13	–0.14890	–0.710
T14	–0.43829	–1.621
T15	–0.49780	–1.376
T16	–0.58153	–1.341

T17	–0.27998	–0.391
T18	–0.29979	–0.975
T19	–0.27616	–0.796
T20	–0.32431	–0.889
T21	–0.32216	–1.327
T22	0.12731	0.534
Driver's class		
CL7	–0.40895	–3.557
CL8	0.47235	1.319
CL9	–0.09367	–0.871
CL10	–3.31830	–0.095
CL11	0.75389	4.824
CL12	0.38643	2.935
CL13	0.19255	2.036
CL18	–0.30438	–1.702
CL19	–0.66526	–4.364
NEW	–0.17552	–1.436
AGEAUTO	0.05828	3.328
Number of observations	4,772	
Log-likelihood	–1,642.626	

Table 4.4
Diversification Level of Studied Portfolios

Number of different assets	Combination of assets	Proportion (%)
0		9.2
1	Liquid Asset	21.6
	HBS	2.4
	Life Insurance	1.5
	Stock and Bond	1.1
	Total	26.6
2	Liquid Asset + HBS	10.2
	Liquid Asset + Life Insurance	7.7
	Liquid Asset + Stock and Bond	7.6
	HBS + Life Insurance	1.2
	HBS + Stock and Bond	0.8
	Stock and Bond + Life Insurance	0.5
	Total	28.0
3	Liquid Asset + HBS + Life Insurance	7.5
	Liquid Asset + Stock and Bond + Life Insurance	5.7
	Liquid Asset + HBS + Stock and Bond	7.8
	HBS + Stock and Bond + Life Insurance	0.8
	Total	22.0
4	Liquid Asset + HBS + Stock and Bond + Life Insurance	12.4

Table 4.5
Estimation of the Logit Model for Life Insurance

Variable	Model 4 Conditional on the decision variables only		Model 5 Conditional on the decision variables and their expectations	
	Coefficient	**Wald Chi-square statistic**	**Coefficient**	**Wald Chi-square statistic**
Intercept	–3.0340	101.1371	–16.1711	444.8785
Age 1	0.5480	28.4901	1.4229	65.9224
Age 2	–0.0610	35.0456	–0.1121	75.2270
Income	0.0134	11.3471	–0.0070	0.8014
Total Wealth	2.5625	347.0983	–0.1809	1.2095
Sex	–0.0510	0.3684	–0.5577	25.6743
Occupation 2	0.1371	1.3144	–0.6870	20.5386
Occupation 3	0.1882	3.3965	–0.4583	14.2203
Occupation 4	0.0799	0.5534	–0.0869	0.3082
Occupation 5	0.0190	0.0378	0.2588	3.1969
Occupation 6	0.2370	3.4280	–0.5786	11.6255
Occupation 7	–0.4840	13.0765	–0.5138	8.4863
District 1	0.2260	8.6343	0.0120	0.0131
District 2	0.1817	5.0111	–0.0205	0.0441
District 3	0.2946	12.0767	–0.0938	0.9024
District 4	0.3225	20.4899	0.0223	0.0783
Education 2	0.0256	0.1500	–0.0776	1.1562
Education 3	0.0725	1.0913	–0.2713	12.6975
Education 4	–0.0613	0.3656	–0.0829	0.4958
Housing 1	0.1946	3.2427	0.0815	0.5018
Housing 2	–0.0424	0.1448	0.5807	18.6138
Household type 2	–0.2743	7.9454	0.8497	41.8249
Household type 3	–0.1452	2.2975	–0.5300	19.6440
Household type 4	–0.0270	0.0610	–0.7857	33.1828
Household type 5	–0.2688	7.2596	–0.3562	10.8407
Household type 6	–0.2687	8.1108	–0.2054	3.9184
Liquid asset	0.3964	38.9024	–0.1484	4.3335
HBS	0.3599	57.4634	–0.1288	6.0106
Stock and bond	0.4665	71.8624	0.1357	5.3426
exp. liquid asset			13.6634	645.5160
exp. HBS			3.9539	20.0459
exp. stock and bond			–1.8813	30.0230
Log Likelihood	–6,161.030		–5,677.380	
Number of observations	10.818		10.818	