is to discard all accidents in which only one automobile was involved. Whenever two cars are involved, a declaration is much more likely to be made in any case.\textsuperscript{14}

C. The Procedures

As explained in the Introduction, we want to test the conditional independence of the choice of better coverage and the occurrence of an accident, where “conditional” means conditional on all variables observed by the insurer. To do this, we shall use two parametric methods and three nonparametric methods.

1. A Pair of Probits

We define here our notation. Let \( i = 1, \ldots, n \) denote individuals. The term \( X_1 \) is the set of exogenous variables for individual \( i \) (these variables will be constants and dummy variables in our application). Also, let \( w_i \) denote the number of 1989 in which individual \( i \) was insured. We now define two 0-1 endogenous variables: (1) \( y_i = 1 \) if \( i \) bought (any form of) comprehensive coverage (a TR contract); \( y_i = 0 \) if \( i \) bought only the minimum legal coverage (an RC contract); (2) \( z_i = 1 \) if \( i \) had at least one accident in which he was judged to be at fault; otherwise (no accident or \( i \) not at fault) it is zero.

These definitions call for two remarks. First, there are many different comprehensive coverage contracts on offer, with (say) different levels of deductible. Ideally, these contracts should be treated separately and not bundled together as we do here. However, this would greatly complicate the model.\textsuperscript{15} Second, we separate accidents in which the insured is at fault and those in which he is not. The reason is that if the insured has an accident in which another driver is to blame, any information on his risk type may not be conveyed. Also, in addition, deductibles are in any case quite small. In expected terms, the difference between the various levels of deductible available in our sample is negligible when compared to that between comprehensive and partial coverage, on which we concentrate here. Experience rating is more important, but, by law, experience rating schemes are identical across firms.

\textsuperscript{14} In principle, the two drivers may agree on some bilateral transfer and thus avoid the penalties arising from experience rating. Such a deal is, however, quite difficult to implement between individuals who meet randomly, will probably never meet again, and cannot commit in any legally enforceable way (since declaration is compulsory according to insurance contracts). We follow the general opinion in the profession that such bilateral agreements can be neglected. An alternative solution is to concentrate on accidents with bodily injuries, since in that case declaration is compulsory by law (see, e.g., Levitt and Porter 1998). However, this drastically reduces the number of accidents.

\textsuperscript{15} In addition, deductibles are in any case quite small. In expected terms, the difference between the various levels of deductible available in our sample is negligible when compared to that between comprehensive and partial coverage, on which we concentrate here. Experience rating is more important, but, by law, experience rating schemes are identical across firms.
we do not exploit the further information linked to drivers who had several accidents in 1989; again, there are very few of these cases.

We now set up two probit models, one for the choice of coverage and one for the occurrence of an accident. Let \( \epsilon_i \) and \( \eta_i \) be two independent centered normal errors with unit variance. Then

\[
y_i = 1(X \beta + \epsilon_i > 0)
\]

and

\[
z_i = 1(X \gamma + \eta_i > 0).
\]

We first estimate these two probits independently, weighing each individual by the number of days under insurance \( w_i \). Then we can easily compute the generalized residuals \( \hat{\epsilon}_i \) and \( \hat{\eta}_i \). For instance, \( \hat{\epsilon}_i \) is given by

\[
\hat{\epsilon}_i = E(\epsilon_i | y_i) = \frac{\Phi(X \beta)}{\Phi(-X \beta)} y_i - (1 - y_i) \frac{\phi(X \beta)}{\Phi(-X \beta)},
\]

where \( \phi \) and \( \Phi \) denote the density and the cumulative distribution function (cdf) of \( N(0, 1) \). Now define a test statistic by

\[
W = \frac{\left( \sum_{i=1}^{n} w_i \hat{\epsilon}_i \hat{\eta}_i \right)^2}{\sum_{i=1}^{n} w_i^2 \hat{\epsilon}_i^2 \hat{\eta}_i^2}.
\]

The general results in Gourieroux et al. (1987) imply that under the null of conditional independence \( \text{cov}(\epsilon_i, \eta_i) = 0 \), \( W \) is distributed asymptotically as a \( \chi^2(1) \). This provides us with a test of the symmetric information assumption.

To implement this procedure, we first need to choose what exogenous variables to include in \( X_i \). The most contentious variable here is the past driving record, as represented by the bonus/malus coefficient (defined above). If we exclude this variable, then we neglect some of the insurer’s information and our test will be biased. If we include it, it may also be biased since this variable is likely to be correlated with \( \eta_i \). As indicated above, our solution is to focus in a first step on drivers who have no past driving record, or “beginners.” In our data, this refers to drivers who obtained their driver’s license in 1988. There are 6,333 of them in our sample. As for the other exogenous variables, we chose to include those that, according to the insurers, are most relevant: thus we have dummy variables for sex (1), make of car (7), performance of the car (5), type of use (3).
2. A Bivariate Probit

Estimating the two probits independently is appropriate under conditional independence, but it is inefficient under the alternative. For this reason, we also estimate a bivariate probit in which \( \varepsilon_i \) and \( \eta_i \) are still distributed as \( N(0, 1) \) but have a correlation coefficient \( \rho \), which we also estimate. This will allow us to test \( \rho = 0 \) but also to get a confidence interval for \( \rho \).

3. A \( \chi^2 \) Test

The two parametric procedures presented above rely on a fairly large number of exogenous variables, as compared to those in Puelz and Snow (1994), for instance. However, the functional forms involved are still relatively restrictive since the latent models are linear and the errors are normal. It is quite possible, for instance, that the data-generating process is driven by cross effects or even more complicated nonlinear functions of the exogenous variables. Then our results would be biased in unpredictable ways. To remedy this, we adopt here a fully nonparametric procedure based on \( \chi^2 \) tests for independence.

Let \( x_i \) be a set of \( m \) exogenous 0-1 variables. Then we can define \( 2^m \) “cells” in which all individuals have the same values for all variables in \( x_i \). In each cell, we compute a two by two table generated by the values of \( y_i \) and \( z_i \), as in Table 1. For \( j = 0, 1 \), define \( N_j = N_{j0} + N_{j1} \), \( N_j = N_{0j} + N_{1j} \), and \( N = N_0 + N_1 \). Now consider the test statistic

\[
T = \sum_{j,k=0,1} \left[ \frac{N_{jk} - (N_j N_k / N)^2}{N_j} \right].
\]


Gouriéroux, Christian; Monfort, Alain; Renault, Eric; and Trognon, Alain. “Generalised Residuals.” J. Econometrics 34 (January/February 1987): 5–32.


