Working document[1]

The moral hazard effect in car insurance: empirical evidence from Ireland

Alexandre Godzinski

February 2009

Research director: Pierre-Yves Geoffard (PSE)

Very preliminary
Confidential

Abstract

This document focuses on identifying moral hazard in car insurance. Theoretically, as monitoring perfectly the effort to drive carefully is impossible, a moral hazard effect should occur. Empirically, it is very difficult to distinguish between adverse selection and moral hazard, known as an empirical puzzle. Following the existing literature in this area, we use an Irish database to test for moral hazard. In a first step we show that information is asymmetric, using the approach based on the cross conditional correlation between coverage and sinistrality. In a second step, we study the presence of moral hazard using particularities of the Irish experience rating system. Under a stationarity assumption, we conclude that there is moral hazard in the Irish data. Without stationarity assumption, our data does not allow us to conclude for the moment.

[1] This document is based on my master thesis, presented at PSE. I would like to thank particularly Pierre-Yves Geoffard (my research director), Philippe Février, Derek Bain and Ben Deans.

Contents

1 Introduction

Insurance theory predicts a transfer of risk from the more risk-averse agents to the less risk-averse. That is why insurees (assumed to be strictly risk-averse) transfer their risks to insurers (assumed to be risk-neutral). In the case of a pure premium, the optimal contract for the insuree is full coverage. However, the larger the share of the risk supported by the insurer, the less the insuree is impacted by a claim, and the less incentive the insuree has to behave carefully. Consequently, a moral hazard effect should be observed (which can be ex ante or ex post). To avoid these hidden actions, claims are made costly to insurees, through the experience rating system (bonus/malus) or deductibles. In France as in Ireland, car insurance uses an experience rating system. Its double mission is to divide the population into classes in which risk is homogeneous, to fight adverse selection, and since they will have to pay more ex post as the premium will increase in the case of a claim, to encourage insurees to drive carefully, to fight moral hazard. This experience rating system (bonus/malus) can be rigidly defined by the law, as in France, or more freedom can be given to the insurer, as in Ireland.

In the case of adverse selection as in the case of moral hazard, a robust prediction derived from Rothschild and Stiglitz (1976) is that the correlation between the level of coverage and the occurrence of claim, conditional on observables, is positive. This prediction, testable from cross data, or static, is tested in Chiappori and Salanie (2000). However, in the case of asymmetry of information, the interesting and difficult point is to distinguish adverse selection from moral hazard. In the case of adverse selection, agents who are more risky choose better coverage. In the case of moral hazard, we have an opposite causality: agents who have better coverage become more risky because more comprehensive coverage lowers incentives to drive carefully. Two recent works study this distinction in car insurance: on the one hand Abbring, Chiappori and Pinquet (2003) and on the other hand Dionne, Michaud and Dahchour (2004). In both cases, longitudinal, or dynamic, data are used. Indeed, finding moral hazard in the presence of unobserved heterogeneities can be seen as a particular case of the distinction between state dependency and pure heterogeneity, a distinction (studied by Heckman) which is possible to make with dynamic data. An older case, presented in Heckman (1981), is to analyze causes making an unemployed person less employable than an employed person, observable characteristics being fixed. A first explanation is a stigma effect (explanation in terms of state dependency). A second one is that an unobserved characteristic makes the unemployed person less able to be employed in the present and in the future (explanation in term of pure heterogeneity).

Recent empirical works tackling the question of testing the presence of asymmetry of information in car insurance leads to opposite conclusions, which must therefore be relativized by differences in available information and in the structure of contracts. With regard to testing for the presence of information asymmetry, Chiappori and Salanie (2000), who use a database provided by the French Federation of Insurance Companies, cannot reject the hypothesis of absence of asymmetry of information, whereas Abbring, Chiappori and Pinquet (2003), who use a database provided by a French insurer, conclude that information asymmetry is present, using the bonus/malus level and mileage. Regarding the test of presence of moral hazard, Abbring, Chiappori and Pinquet (2003), who use a database provided by a French insurer, cannot reject the hypothesis of absence of moral hazard, whereas Dionne, Michaud and Dahchour (2004) conclude that there is moral hazard in their data. Using a longitudinal database provided by an Irish insurer, the purpose of this work is to test first of all the presence of asymmetry of information and secondly the presence of moral hazard, contributing to the literature related to this empirical puzzle.

Part 2 presents a simple framework of moral hazard in insurance. Part 3 summarizes existing works in the field of testing the presence of asymmetry of information in car insurance and in health insurance. Part 4 presents the Irish database, the available information, the contract menu and the structure of the experience rating system used by the Irish insurer. Some descriptive statistics are presented, permitting an initial understanding of the situation useful in implementing tests in the following parts. Part 5 tests the presence of asymmetry of information using the cross conditional correlation between coverage and sinistrality property for different times. Part 6 tests the presence of moral hazard. Section 6.1 presents the strategy for testing. This consisting of using the fact that the experience rated system used by the Irish insurance provides only a bonus and no malus, and consequently, insurees who have no bonus see their incentives to drive carefully lowered after a first claim because they know that they will not get a bonus the following year. After the contract is renewed (during the second year) however, they can hope to get a discount for the third year, and consequently have more incentive to drive carefully than at the end of the first year. Section 6.2 focuses on the incentive change when the contract is renewed, seen as an exogenous change. Section 6.3 focuses on the incentive change before and after the first claim, seen as an endogenous change. The endogenous characteristic of the change leads to the need for a more complex model, where the analysis and the tests of Abbring, Chiappori and Pinquet are reused and adapted.

2 Theoretical framework

In this part a simple framework focusing on the probability of claim under different assumptions is presented.

In order to have a simple but structured view of factors influencing the probability of claim, let us consider an individual facing a unique risk, which he at least partially bears, and whose probability of claim is noted p. Factors determining this probability, under the assumption of a homogeneous population or of a heterogeneous population whose parameter is λ (heterogeneity which can come from a difference in the level of risk or in the risk aversion), and under the assumption that the individual can impact his probability of claim by choosing the optimal effort e^* (presence of moral hazard) or not (absence of moral hazard), are summarized in the following table:

	Homogeneous population	Heterogeneous population
Without moral hazard	p constant	$p(\lambda)$
With moral hazard	$p(e^*)$	$p(\lambda, e^*)$

Thus, with the exception of the case without moral hazard and in which the population is homogeneous, the probability p is distributed under a distribution which is not a Dirac. In a plausible framework, the population is heterogeneous (because even by conditioning by the observable information, it is likely that unobservable heterogeneities are still present), and consequently the potential impact of effort is disturbed by residual heterogeneity, which gives an idea of the difficulty of studying the impact of an incentive effect in the presence of selection effects.

However, let us try to present a simple model taking into account moral hazard and in which the population is heterogeneous. Let us study the impact of the individual's state, noted s, on the probability of claim p. This state s depends in the more general case both on choices of the individual (coverage and effort choice) and on the realization of past shocks on the sinistrality. The utility function of the individual depends on three parameters: his type λ, the state of his insurance coverage s and his effort e:

$$U(\lambda, s, e)$$

The individual maximizes his expected utility $\mathbb{E}U(\lambda, s, e)$. A more complete framework would require a multi period model, for instance by using a Bellman equation, as it is in Abbring, Chiappori and Pinquet (2003), where the effort is continuous, or by adapting a model derived by Shapiro and Stiglitz (1984), where the effort is discrete, but we can reasonably assume that the expected utility takes the form of an expected profit $\mathbb{E}\pi(\lambda, s, e)$ minus the effort cost $C(e, \lambda)$. The individual chooses his effort to maximize his expected utility, hence:

$$e^*(\lambda, s) = \arg\max_e (\mathbb{E}\pi(\lambda, s, e) - C(e, \lambda))$$

Let us express the type λ of the individual as a function of the observed information x and the unobserved information θ: $\lambda = f(x, \theta)$.

Hence the probability of claim is $p(\lambda, e^*(\lambda, s)) = p(\lambda, s) = p(x, \theta, s)$.

As indicated previously, with conditioning by the observed information x, it is possible to work with $p(x, \theta, s)|x = p(\theta, s)$, but there is always a heterogeneity in the sinistrality, and the question lies in being able to separate what is related to an unobserved heterogeneity (through θ) from what is related to a state dependency (through s).

Intuitively, concerning the impact of the state s for a given individual, the more the state leads to an important risk transfer towards the insurer, the weaker the effort is and the higher the probability of claim. This result can be found, assuming that the maximization occurs in good conditions: $\frac{\partial \mathbb{E}U}{\partial e}(\lambda, s, e^*) = 0$ (maximum reached in an interior point) and $\frac{\partial^2 \mathbb{E}U}{\partial e^2}(\lambda, s, e^*) < 0$ (expected utility concave in effort), that the better the state, the less the increase in effort increases the expected profit $\frac{\partial^2 \mathbb{E}\pi}{\partial e \partial s} < 0$ (increasing the effort impacts less and less the expected profit when more risk is transferred to the insurer), that the effort cost is increasing in the effort: $\frac{\partial C}{\partial e}(e, \lambda) > 0$ and that the probability of claim is decreasing in the effort.

Let us write the first order condition:

$$\frac{\partial \mathbb{E}U}{\partial e}(\lambda, s, e^*(\lambda, s)) = 0$$

And let us derive with respect to the state s:

Hence:

$$\frac{\partial^2 \mathbb{E}U}{\partial s \partial e}(\lambda, s, e^*(\lambda, s)) + \frac{\partial e^*}{\partial s}(\lambda, s) \frac{\partial^2 \mathbb{E}U}{\partial e^2}(\lambda, s, e^*(\lambda, s)) = 0$$

$$\frac{\partial e^*}{\partial s}(\lambda, s) \underbrace{\frac{\partial^2 \mathbb{E}U}{\partial e^2}(\lambda, s, e^*(\lambda, s))}_{<0} = \underbrace{-\frac{\partial^2 \pi}{\partial e \partial s}(\lambda, s, e^*(\lambda, s))}_{>0}$$

Thus:

$$\frac{\partial e^*}{\partial s}(\lambda, s) < 0$$

So in accordance with the intuition, the better the state s, the lower the effort, and the higher the occurrence of a claim.

Let us consider an individual of type λ in two states s^1 and s^2 such that the state s^1 leads to a better coverage than the state s^2. We can deduce from below that $p(\lambda, s^1) > p(\lambda, s^2)$. On the other hand, in a model without moral hazard, the probability p does not depend on the state. Consequently, all else equals, a change in the state has no impact on the probability p.

Therefore the moral hazard effect implies a positive partial correlation between the state of the coverage and the sinistrality, but this positive partial correlation is not a characteristic of moral hazard. In fact, a pure adverse selection effect, by reverse causality, leads to the same prediction. For memory, these effects are summarized in the following table:

Pure adverse selection	$risk\ \lambda \overset{contrat\_choice}{\Rightarrow} state\ of\ coverage$
Pure moral hazard	$state\ of\ coverage \overset{effort\_choice}{\Rightarrow} level\ of\ risk\ p(\lambda, e)$

3 Literature review

Several works focused on the question of the test of the presence of information asymmetry and of moral hazard. In the field of car insurance, we can quote the works of Puelz and Snow (1994), Dionne and Vanasse (1992), Chiappori and Salanie (1997) and (2000), Dionne, Gourieroux, and Vanaase (1999) and (2001), Richaudeau (1999), and, in the field of health insurance and life insurance those of Hollym Gardiol, Domeninghetti, and Bisig (1998), Chiappori, Durand, and Geoffard (1998), Chiappori, Geoffard, and Kyridizou (1998), Cardon and Hendel (1998), Hendel and Lizzeri (1999). Some works focused on distinguishing empirically adverse selection from moral hazard, for instance Holly et al. (1998) and Cardon and Hendel (1998) who use structural models in health insurance, or Chiappori, Durand, Geoffard (1998) and Dionne, Maurice, Pinquet, and Vanasse (2001) who use natural experiments in which new legislation exogenously changes the incentives.

Concerning the question of distinguishing adverse selection and moral hazard, Gardiol, Geoffard and Grandchamp (2005) focus on the question of quantitatively distinguishing selection effects from incentive effect using Swiss health insurance data, and conclude that between these two effects, 75% can be attributed to selection effects and 25% to incentive effects.

However, only concerning the car insurance in France, empirical conclusions are not unanimous. In recent works, Abbring, Chiappori, and Pinquet (2003), whose idea is summarized above, conclude the absence of moral hazard (more precisely, their data do not allow them to reject the null hypothesis of absence of moral hazard), whereas Dionne, Michaud, and Dahchour (2004), using a Granger causality test, conclude the presence of moral hazard in their data.

Empirically showing the presence of moral hazard is usually done in two steps. In a first step, we study the correlation between coverage and sinistrality, controlling the other observable characteristics, which allows us to conclude the presence of information asymmetry in the case of positive partial correlation, which is for instance studied in Chiappori and Salanie (1997) and Chiappori, Jullien, Salanie, and Salanie (2001). In a second step, in the case of presence of information asymmetry, we have to distinguish between what comes from moral hazard and what comes from adverse selection, and on this point, much more delicate, longitudinal, dynamic, data represents a strong advantage compared to cross, static data. One way to proceed is for example developed in Abbring, Chiappori and Pinquet (2003) and in Abbring, Chiappori, Heckman, and Pinquet (2003): in the presence of moral hazard, the probability of having a claim, conditionally on observable parameters and on unobservable heterogeneity, should discontinuously decrease in the case of a claim, whereas without moral hazard, this probability to have a claim should not be affected by a claim. This result comes from the fact that the marginal cost of a claim in France increases with the number of past claims (result found by using the multiplicative character of the malus in France), a new claim hence increasing incentives to drive carefully.

From the econometrics point of view, an underlying question is the distinction between pure heterogeneity and state dependency. Showing the presence of moral hazard can be seen as studying state dependency in the presence of unobservable heterogeneity (in the case of moral hazard, the coverage, considered as given, leads to the choice of effort, which impacts sinistrality, hence state dependency. In the case of adverse selection, however, individual type, unobservable heterogeneity, leads to the choice the level of coverage). And to distinguish pure heterogeneity from state dependency, dynamic data are required. This is based on Heckman's work (for instance Heckman (1981)).

4 The Irish database

The dataset provided by a major Irish insurer allows us to follow monthly characteristics and claims of more that a million insurees. The observation time ranges from January 2004 to March 2008, which means that it is possible to follow an insuree at most during 51 months. This descriptive part is needed to establish tests in the two following parts.

4.1 Available information, contract menu and no claim discount mechanism

The raw database has 28 variables and at least one record per insuree and per month, which give it the size of an area of 35,828,632 rows and 28 columns. Beside variables related to the time of the record, information provided by the insurer can be divided into four categories:

- Information related to characteristics of the insuree: age, gender, geographic area, driving licence type (provisional or not) and number of additional drivers.
- Information related to characteristics of the vehicle: vehicle age, year of record, category, use.
- Information related to characteristics of the contract, linked to insuree's choices and past sinistrality: chosen contract type, last renewal time, last premium paid, company which has sold the contract, bonus level and bonus protection type.
- Information related to the sinistrality of the month of record. The monthly sinistrality is divided into four categories if it is related to damages of the insuree vehicle, to third party damages, to third party injuries, or to other damages. For each category two variables are available, the first counting the claim number within the month and the second giving the aggregate amount in euros within the month.

The contract menu offers three possibilities:

- Third party coverage without fire and theft.
- Third party with fire and theft.
- Comprehensive coverage.

In Ireland, the functioning of the experience rating system is not defined by law but gives the insurer freedom of choice. In the experience rated system used by our major Irish insurer, the bonus level increases in case of no claim and decreases in case of claim, but compared to France the three major differences are: firstly the discount coefficient can only lead to a bonus and not to a malus (the premium paid by the insuree is always equal to or lower than the base premium), secondly the change in the discount coefficient is not obtained by multiplication by a coefficient and thirdly the discount level has less possible values. The evolution of the discount level under the past sinistrality is given by the following table:

Number of claim free year	Discount (%)	Discount after third party claim (%)	Discount after accidental damage claim to own car (%)
0	0	0	0
1	20	0	0
2	30	0	20
3	40	0	30
4	50	20	40
$\geq$5	70	30	50

Another particularity of the experience rating system used by our major Irish insurer is the existence of a protection of the bonus level. Such a protection leads to no change in the bonus level in case of claim. The state of the protection has five modalities, three related to a state in which the bonus level is protected and two related to a state in which the bonus level is not protected. These five modalities are the following:

- L : Lifetime bonus, granted for free to insurees who are clients of our major Irish insurer for more than 10 years and who report no claim for 10 years (except the case in which the claim is not taken into account because of another protection of the bonus level). The insuree who has a lifetime bonus keeps a 70% discount even in the case of a reported claim. This lifetime bonus can be suppressed if the insuree is convicted over a serious motor offense.
- Y : Purchased protection, for one year. This protection is proposed to clients who reported no claims in three years and its price represents 12.5% of the premium.
- T : Trump card, which is free protection granted for experimental purpose by our major Irish insurer to some clients when they renew their contract. This protection is in force until the next claim and is not renewable.
- N : No protection purchased.
- E : Excluded. Some clients are excluded from the protection system because of their too high past sinistrality.

The state of the bonus can consequently be seen as the pairing of the discount level and the type of protection of the discount level.

Some first treatments were done to the raw database to be able to use it. Observations were aggregated on a monthly level (in case of a change in the characteristics within a month, several records are present for the same month, but the absence of the day of change does not allow us to use them). Only cars owned by individuals (which excludes vehicles owned by companies) for private use were kept. After discussion with our major Irish insurer, the minimum age to drive a car in Ireland is 17, which leaded to a treatment. The retreated database has 24,643,635 observations, allowing us to follow 1,063,695 clients.

The database has extremely few empty observations. Only two variables have missing observations, the vehicle record year and the vehicle group, but these missing observations represent less than 0.1% of the total. The geographic area has the modality "unknown" in one fifth of the observations. Except these three variables, all the others are fully filled.

4.2 Descriptive statistics

Simple statistics regarding quantitative variables of the retreated database, related to insurees and weighted by the length of observation in entire month, are given in the following table:

Variable	Mean	Standard deviation	Minimum	P1	P5	Median	P95	P99	Maximum
Age	43.78	14.57	17	19	23	42	71	81	99
Nb of additional drivers	0.76	0.69	0	0	0	1	2	3	9
Vehicle age	5.91	4.24	0	0	0	5	14	20	20
Vehicule registration year	2000	4.12	1944	1989	1992	2000	2006	2007	2008
Last premium charged	720	500	0	220	291	596	1,636	2,582	300,242
Discount	52	28	0	0	0	70	70	70	70
Nb of AD claims	0.00462	0.06871	0	0	0	0	0	0	11
Nb of TPD claims	0.00266	0.05204	0	0	0	0	0	0	11
Nb of TPI claims	0.00067	0.02609	0	0	0	0	0	0	3
Nb of other claims	0.00269	0.05237	0	0	0	0	0	0	4
Amount of AD claims	7	214	-12,562	0	0	0	0	0	95,467
Amount of TPD claims	4	171	-21,050	0	0	0	0	0	166,739
Amount of TPI claims	13	2,396	-11,087	0	0	0	0	0	5,757,109
Amount of other claims	1	19	-5,909	0	0	0	0	0	5,080

This first table raises several points. More than half of the insurees (65.18%) have the maximum discount of 70%. 20.14% have no discount and the remaining 14.68% are situated between these two extreme cases. Concerning the age of the insuree and of the vehicle, actual values are truncated respectively to 99 years and 20 years. Concerning amounts (in euros), minimum negative values are due to rare cases in which the indemnization system between insurers leads to a net profit for the insurer in case of a claim. The higher maximum amount among the four categories corresponds to a claim leading to injuries, financial implications of which are usually more important than when the claim leads to material damage only.

Regarding qualitative variables, also weighted by the length of observation, 49.16% of clients are men, the complement being formed by women because vehicles owned by companies were not taken into consideration. Only vehicles for private use were taken into consideration. 10.70% of clients have a provisional driving licence. 75.89% of clients have a comprehensive coverage, others having a third party coverage. 34.72% of insurees did not declare any other driver, 64.41% declared one and the 7.87% remaining declared between 2 and 9 additional drivers. 5.95% have a lifetime bonus, 13.99% have purchased a bonus protection, 0.82% have a Trump card, 3.03% are excluded from the bonus protection system because of a too high sinistrality and 76.21% have no bonus protection because they do not meet the length requirement or because they choose not to purchase a bonus protection.

4.3 Stationarity of sinistrality

In order to implement a test of presence of moral hazard, it is useful to focus on the question of stationarity. The two following graphs show the mean of the indicator of having reported at least one claim in the month which implies respectively third party damages and third party injuries.

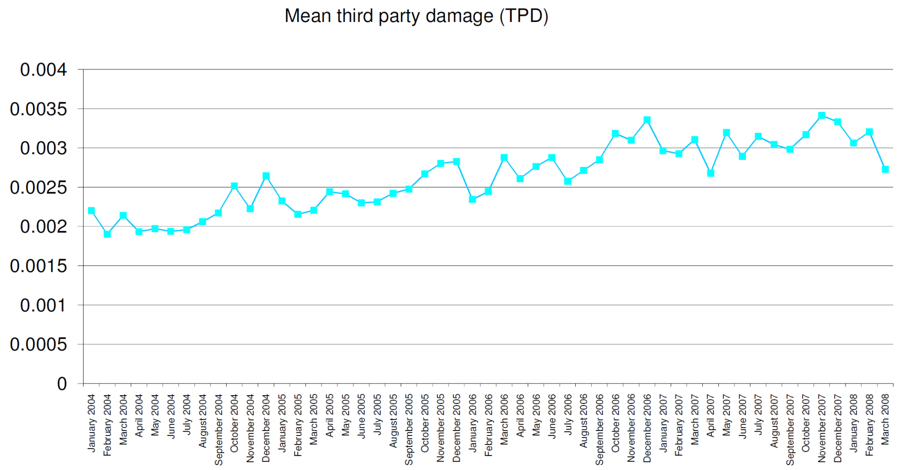

Figure 1: Mean of the indicator to report in the month at least one claim implying third party damage (TPD).

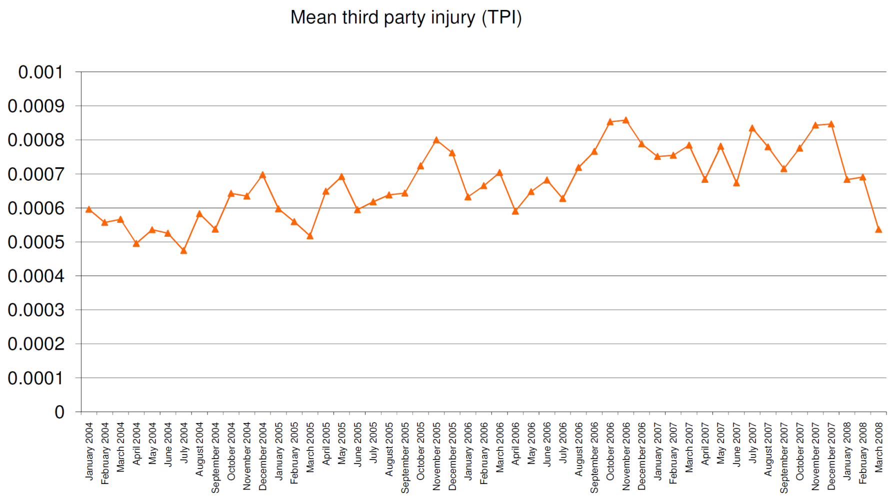

Figure 2: Mean of the indicator to report in the month at least one claim implying third party injury (TPI).

These two graphs raises three points. Firstly, we can see an increasing trend in both of them. There are also some peaks in November, and during the winter, and also usually in May, indicating a seasonal impact. Concerning the increasing trend, it is not possible at this stage to distinguish if the increase comes from the fact that on average a given individual becomes more risky or if it comes from the fact that new entrants in the panel are more risky than individuals exiting the panel. Secondly, claims leading to injuries seem more volatile within time than those leading to damages only. Thirdly, we can see a decrease among the last available months, indicating that some claims are incurred but not yet reported, this phenomenon being more important for third party injury, since developments in this case are likely to be longer.

5 Testing for information asymmetry

5.1 Methodology

In the case of the presence of adverse selection as in the case of the presence of moral hazard, a robust prediction, presented in Chiappori and Salanie (1997) and in Chiappori, Jullien, Salanie and Salanie (2001) is that the correlation between the contract choice (which defines the coverage level) and the claim occurrence, conditionally to observables, is positive. In the case of the presence of adverse selection, agents know if their probability of claim exceeds the average probability in their risk class proposed by the insuree, and in this case will choose better coverage and have more claims. Consequently, conditionally to observables, the choice of a comprehensive coverage is positively correlated with a high level of claim. In the case of the presence of moral hazard, the causality is reverse, agents, who for a certain reason, buy more coverage become riskier because a more comprehensive coverage has a negative impact on incentives and discourages prudent behaviour. Consequently the presence of information asymmetry, which can be due to an adverse selection or moral hazard effect, is testable from this partial correlation. Besides, this prediction is realizable from static data.

Let us consider on the one hand the indicator of choosing a comprehensive coverage (if the indictor is equal to zero, it means that the agent chose a third party coverage, with or without fire and theft included) and on the other hand the indicator of having at least one claim implying an injury (this variable seems less affected by the type of contract or problem of reporting). After a logit regression between this two indicators on the observables, Pearson residuals are considered. The correlation between these two residuals are studied on the one hand parametrically from the linear correlation r (Pearson) and nonparametrically from the rank correlation, using the Spearman rank correlation coefficient ρ_S and the Kendall coefficient τ_b.

The double regression is the following:

$$\begin{cases} n_i = 1(x_i\beta + \epsilon_i^n) > 0 \\ c_i = 1(x_i\gamma + \epsilon_i^c) > 0 \end{cases}$$

For memory, the definition of the linear correlation between two random variables X and Y is the following

$$r = \frac{cov(X, Y)}{\sigma_X . \sigma_Y}$$

Considering ranks of the random variables, the Kendall coefficient τ_b is defined by:

$$\tau_b = \frac{C - D}{\sqrt{(n^2 - E_X)(n^2 - E_Y)}}$$

where n^2 is the total number of pairs of individuals (i, j), C is the number of concordant (i, j) pairs, which means that i and j are in the same order than X and Y, D is the number of discordant pairs, E_X is the number of ex aequo for X, such that $X_i = X_j$, and E_Y is the number of ex aequo for Y. The Spearman rank correlation coefficient ρ_S is the linear correlation coefficient between $rank\ X$ and $rank\ Y$.

5.2 Partial correlation between sinistrality and coverage

The partial correlation between sinistrality and coverage was computed using as available information only characteristics declared by the insuree and then using also the level of bonus, which consequently linked the result to the way the insurer takes into account the past sinistrality.

The developed form of the double regression without using the level of bonus is the following:

$$\begin{cases} sinTPI_i = 1(\beta_0 + \beta_1 HOMME_i + \beta_2 AGE_i + \beta_3 AGE_i^2 + \beta_4 AGE_i^3 \\ +\beta_5 HOMME_i * AGE_i + \beta_6 HOMME_i * AGE_i^2 + \beta_7 HOMME_i * AGE_i^3 \\ +\beta_8 PERMISPROB_i + \beta_9 DUBLIN_i + \beta_{10} AUTRECOND_i \\ +\beta_{11} VEHAGE_i + \beta_{12} VEHIMM_i + \beta_{13} VEHGROUPL_i + \beta_{14} COMP1_i + \beta_{15} COMP2_i + \epsilon_i^n) > 0 \\ \\ couvcomp_i = 1(\gamma_0 + \gamma_1 HOMME_i + \gamma_2 AGE_i + \gamma_3 AGE_i^2 + \gamma_4 AGE_i^3 \\ +\gamma_5 HOMME_i * AGE_i + \gamma_6 HOMME_i * AGE_i^2 + \gamma_7 HOMME_i * AGE_i^3 \\ +\gamma_8 PERMISPROB_i + \gamma_9 DUBLIN_i + \gamma_{10} AUTRECOND_i \\ +\gamma_{11} VEHAGE_i + \gamma_{12} VEHIMM_i + \gamma_{13} VEHGROUPL_i + \gamma_{14} COMP1_i + \gamma_{15} COMP2_i + \epsilon_i^c) > 0 \end{cases}$$

Using the level of bonus, this double regression becomes:

$$\begin{cases} sinTPI_i = 1(\beta_0 + \beta_1 HOMME_i + \beta_2 AGE_i + \beta_3 AGE_i^2 + \beta_4 AGE_i^3 \\ +\beta_5 HOMME_i * AGE_i + \beta_6 HOMME_i * AGE_i^2 + \beta_7 HOMME_i * AGE_i^3 \\ +\beta_8 PERMISPROB_i + \beta_9 DUBLIN_i + \beta_{10} AUTRECOND_i \\ +\beta_{11} VEHAGE_i + \beta_{12} VEHIMM_i + \beta_{13} VEHGROUPL_i + \beta_{14} COMP1_i + \beta_{15} COMP2_i \\ +\beta_{15} BONUSMIN + +\beta_{16} BONUSMAX + \epsilon_i^n) > 0 \\ \\ couvcomp_i = 1(\gamma_0 + \gamma_1 HOMME_i + \gamma_2 AGE_i + \gamma_3 AGE_i^2 + \gamma_4 AGE_i^3 \\ +\gamma_5 HOMME_i * AGE_i + \gamma_6 HOMME_i * AGE_i^2 + \gamma_7 HOMME_i * AGE_i^3 \\ +\gamma_8 PERMISPROB_i + \gamma_9 DUBLIN_i + \gamma_{10} AUTRECOND_i \\ +\gamma_{11} VEHAGE_i + \gamma_{12} VEHIMM_i + \gamma_{13} VEHGROUPL_i + \gamma_{14} COMP1_i + \gamma_{15} COMP2_i \\ +\gamma_{15} BONUSMIN + +\gamma_{16} BONUSMAX + \epsilon_i^c) > 0 \end{cases}$$

In order to be easily readable, explanatory variables are summarized in the following table:

Variable explicative	Signification
HOMME	Gender
AGE	Age
AGE^2	Age square
AGE^3	Age cube
HOMME*AGE	Gender crossed with age
$HOMME*AGE^2$	Gender crossed with age square
$HOMME*AGE^3$	Gender crossed with age cube
PERMISPROB	Provisional license
DUBLIN	Insuree living in Dublin
AUTRECOND	Existence of at least one other driver
VEHAGE	Vehicle age
VEHIMM	Vehicle Registration Year
VEHGROUPL	Low risk vehicule
COMP1	Contract sold by company 1
COMP2	Contract sold by company 2
BONUSMIN	Minimum discount (0%)
BONUSMAX	Maximum discount (70%)

The following table summarizes results, presenting the linear correlation coefficient r, the Spearman rank correlation coefficient ρ_S and the Kendall τ_b coefficient, each time seconded with their p-value to be able to judge the pertinence of the hypothesis of a null coefficient. For each month, from January 2004 to March 2008, results are presented when the past sinistrality is not taken into consideration. In the second part of the table, yearly aggregate data were used, keeping only individuals present during the twelve months of the considered year. Results with and without taking account of the past sinistrality are presented.

Date	Nb obs	logit sin	logit cov	Correlation coefficient					
		-2 log L	-2 log L	Pearson r	p-value	Spearman ρ_S	p-value	Kendall τ_b	p-value
January 2004	478 159	4 802	566 368	0.00162	0.2620	0.07883***	< 0.0001	0.06881***	< 0.0001
February 2004	479 023	4 535	567 294	-0.00060	0.6795	0.09109***	< 0.0001	0.06621***	< 0.0001
March 2004	478 273	4 594	566 091	0.00018	0.8988	0.03197***	< 0.0001	0.02712***	< 0.0001
April 2004	476 112	4 064	562 908	0.00027	0.8543	0.06336***	< 0.0001	0.04652***	< 0.0001
May 2004	475 818	4 351	561 687	-0.00070	0.6278	0.05033***	< 0.0001	0.04063***	< 0.0001
June 2004	477 663	4 293	561 751	0.00020	0.8927	0.08213***	< 0.0001	0.04967***	< 0.0001
July 2004	482 251	3 963	565 125	0.00196	0.1743	0.05432***	< 0.0001	0.02925***	< 0.0001
August 2004	486 921	4 798	567 837	0.00124	0.3866	0.06140***	< 0.0001	0.03450***	< 0.0001
September 2004	491 132	4 503	570 533	-0.00049	0.7327	0.09342***	< 0.0001	0.05222***	< 0.0001
October 2004	493 054	5 293	570 389	-0.00080	0.5746	0.10301***	< 0.0001	0.08900***	< 0.0001
November 2004	493 103	5 220	568 654	0.00037	0.7940	0.11435***	< 0.0001	0.09377***	< 0.0001
December 2004	490 190	5 655	564 279	-0.00074	0.6032	0.10674***	< 0.0001	0.05931***	< 0.0001
January 2005	491 980	4 952	564 500	0.00052	0.7148	0.07979***	< 0.0001	0.05678***	< 0.0001
February 2005	491 443	4 668	562 518	-0.00128	0.3677	0.08496***	< 0.0001	0.06065***	< 0.0001
March 2005	490 357	4 351	559 437	-0.00124	0.3861	0.10148***	< 0.0001	0.07227***	< 0.0001
April 2005	489 951	5 304	557 488	0.00107	0.4538	0.08835***	< 0.0001	0.05878***	< 0.0001
May 2005	489 738	5 611	555 402	-0.00092	0.5202	0.06533***	< 0.0001	0.04525***	< 0.0001
June 2005	487 686	4 888	551 154	0.00154	0.2813	0.10200***	< 0.0001	0.07405***	< 0.0001
July 2005	485 333	5 033	546 891	-0.00034	0.8125	0.03690***	< 0.0001	0.02900***	< 0.0001
August 2005	482 547	5 148	541 718	0.00025	0.8612	0.11484***	< 0.0001	0.08719***	< 0.0001
September 2005	480 135	5 159	537 241	-0.00073	0.6113	0.06653***	< 0.0001	0.03013***	< 0.0001
October 2005	476 525	5 664	531 193	0.00021	0.8827	0.06066***	< 0.0001	0.04314***	< 0.0001
November 2005	473 658	6 163	526 476	0.00022	0.8809	0.10939***	< 0.0001	0.08446***	< 0.0001
December 2005	470 988	5 872	522 108	-0.00115	0.4294	0.06385***	< 0.0001	0.04078***	< 0.0001
January 2006	475 809	5 036	524 855	0.00123	0.3975	0.12820***	< 0.0001	0.10299***	< 0.0001
February 2006	479 270	5 290	526 694	0.00008	0.9577	0.12012***	< 0.0001	0.07071***	< 0.0001
March 2006	482 858	5 616	529 129	0.00112	0.4350	0.09863***	< 0.0001	0.07015***	< 0.0001
April 2006	484 381	4 824	528 985	0.00016	0.9133	0.06870***	< 0.0001	0.04645***	< 0.0001
May 2006	486 026	5 240	529 522	0.00276*	0.0548	0.13244***	< 0.0001	0.08941***	< 0.0001
June 2006	486 772	5 505	528 897	-0.00071	0.6200	0.06121***	< 0.0001	0.04302***	< 0.0001
July 2006	485 683	5 107	526 672	0.00147	0.3050	0.05539***	< 0.0001	0.03677***	< 0.0001
August 2006	483 858	5 733	523 647	-0.00192	0.1824	0.09563***	< 0.0001	0.06301***	< 0.0001
September 2006	481 598	6 018	520 269	-0.00075	0.6018	0.10907***	< 0.0001	0.08421***	< 0.0001
October 2006	479 129	6 598	516 796	-0.00124	0.3925	0.08843***	< 0.0001	0.06158***	< 0.0001
November 2006	476 388	6 565	513 093	-0.00253*	0.0809	0.07732***	< 0.0001	0.04562***	< 0.0001
December 2006	471 805	6 060	507 230	-0.00110	0.4489	0.13529***	< 0.0001	0.09510***	< 0.0001
January 2007	473 668	5 833	507 568	0.00128	0.3795	0.09851***	< 0.0001	0.07412***	< 0.0001
February 2007	472 935	5 847	505 209	-0.00070	0.6278	0.12613***	< 0.0001	0.09602***	< 0.0001
March 2007	473 965	6 049	504 043	-0.00026	0.8585	0.07809***	< 0.0001	0.05060***	< 0.0001
April 2007	475 018	5 387	502 772	-0.00033	0.8176	0.12612***	< 0.0001	0.08561***	< 0.0001
May 2007	477 248	6 068	502 646	-0.00048	0.7395	0.10741***	< 0.0001	0.07063***	< 0.0001
June 2007	479 502	5 363	502 130	-0.00002	0.9908	0.06189***	< 0.0001	0.04203***	< 0.0001
July 2007	482 602	6 519	502 764	0.00065	0.6540	0.10189***	< 0.0001	0.06629***	< 0.0001
August 2007	486 079	6 168	502 747	0.00094	0.5119	0.12252***	< 0.0001	0.08682***	< 0.0001
September 2007	489 027	5 769	503 221	0.00221	0.1215	0.09607***	< 0.0001	0.06294***	< 0.0001
October 2007	491 076	6 219	503 916	0.00025	0.8636	0.01057***	< 0.0001	0.00625***	< 0.0001
November 2007	490 762	6 674	502 071	0.00180	0.2068	0.12140***	< 0.0001	0.09004***	< 0.0001
December 2007	487 615	6 669	497 505	-0.00296**	0.0385	0.08068***	< 0.0001	0.05972***	< 0.0001
January 2008	491 514	5 569	498 996	0.00008	0.9565	0.08738***	< 0.0001	0.05769***	< 0.0001
February 2008	493 429	5 630	498 629	0.00103	0.4709	0.06545***	< 0.0001	0.04360***	< 0.0001
March 2008	489 925	4 487	492 444	0.00186	0.1922	-0.02933***	< 0.0001	-0.02082***	< 0.0001
2004[1]	366 182	28 837	419 319	0.00270	0.1021	0.09176***	< 0.0001	0.07356***	< 0.0001
2005[1]	378 735	32 431	412 245	0.00151	0.3532	0.12103***	< 0.0001	0.09702***	< 0.0001
2006[1]	373 608	33 334	387 006	0.00024	0.8846	0.16909***	< 0.0001	0.13337***	< 0.0001
2007[1]	355 175	33 008	351 426	0.00552***	0.0010	0.20084***	< 0.0001	0.16282***	< 0.0001
2004[1,2]	366 182	28 837	419 319	0.00319*	0.0539	0.07381***	< 0.0001	0.06414***	< 0.0001
2005[1,2]	378 735	32 431	412 245	0.00160	0.3252	0.07034***	< 0.0001	0.07153***	< 0.0001
2006[1,2]	373 608	33 334	387 006	-0.00019	0.9063	0.11718***	< 0.0001	0.10852***	< 0.0001
2007[1,2]	355 175	33 008	351 426	0.00551***	0.0010	0.16033***	< 0.0001	0.16513***	< 0.0001

Significance at 1% level are indicated by ***, at 5% by ** and at 10% by *.

The linear correlation on monthly data is almost never significant, even at 10%. The three exceptions are May 2006, November 2006 and December 2007. However, the Spearman rank correlation coefficient ρ_S and the Kendall coefficient τ_b are always very significant. On yearly aggregated data, year 2005 and

[1] Insurees present in the panel during the 12 months of the year considered.
[2] With taking into account the past sinistrality through the discount level.

2006 do not allow us to conclude, but the year 2007, and in a weaker way 2004 when the past sinistrality is taken into account, allows us to reject the null hypothesis of absence of information asymmetry.

Concerning the monthly data, coefficients from the work on ranks are represented in the following figure:

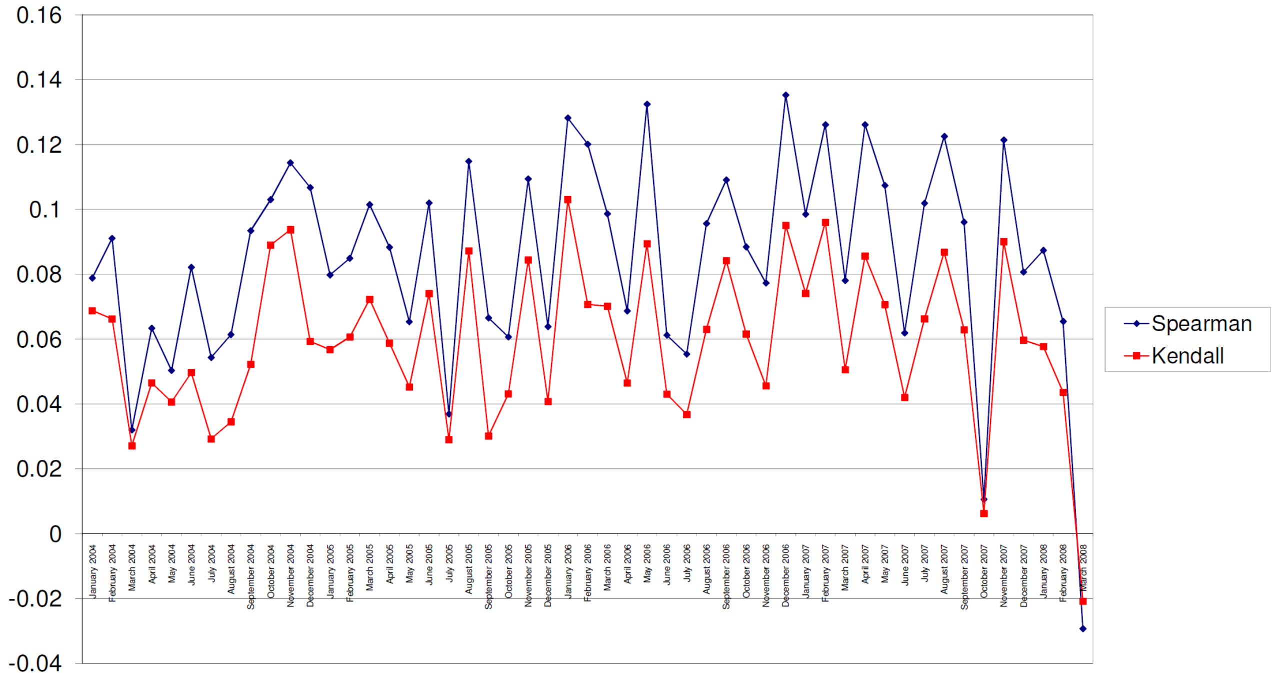

Figure 3: Spearman ρ_S rank correlation coefficient and Kendall τ_b coefficient as a function of time.

We see a clear change in the behaviour in the last months, change which was felt in section 5.3 when studying the stationarity of sinistrality. The presence of incurred but not reported claims can be an explanation, so in order not to pollute results, observation related to the year 2008 will not be used in the following part.

6 Testing for moral hazard

Part 3 allowed us to conclude to the presence of information asymmetry, which legitimates the focus on the presence of moral hazard in this part.

6.1 Strategy

One of the particularities of the experience rating system used by our Irish insurer is that he offers a bonus and no malus (the coefficient by which is multiplied the base premium to obtain the premium paid by the insuree is always lower or equal to one, never strictly greater). Let us consider an insuree with no bonus. He has incentive to drive carefully because if he reports no claim in the year, his discount will switch the following year from 0% to 20%, according to the table provided in section 4.1. Let us assume that this insuree reports a claim during the first year. His incentives to drive carefully instantaneously decrease because he knows that his discount will stay at 0% the following year, whatever the number of additional claims he declares during the end of the first year. However, during the following year (the second year), he can hope to have no claim and to get a discount for the third year. When the contract is renewed, incentives to drive carefully consequently increase instantaneously. The whole strategy on how to identify moral hazard in this work consists of studying the impact on the sinistrality of these two changes of incentives, which is made possible by the monthly following of individuals.

Formally, let us follow an insuree with no bonus during two years and let us assume that he faces a claim during the first year. Let us call period A the period which starts at the beginning of the first year and ends when the first claim occurs, period B the period which starts when the first claim occurs and ends at the end of the first year, and period C the second year. This time partition is summarized by the following scheme:

$$(\overbrace{\underbrace{0-0-1}_{Period\ A}\ \underbrace{0-1-0}_{Period\ B}}^{1st\ year};\overbrace{\underbrace{0-0-0-0}_{Period\ C}}^{2nd\ year})$$

The simplest change to study is the switch from period B to period C, because the moment of the change is the same for each insuree and does not depend on the sinistrality. We will consider this change as exogenous. Tests using the incentives change between periods B and C are presented in section 6.2.

The incentives change between periods A and B is also interesting but more delicate to study, because it is not possible to directly study changes in the level of sinistrality, since we imposed conditions on what happens during period A to select insurees. The direct comparison of the level of sinistrality is thus irrelevant. We will reuse and adapt the model and tests of Abbring, Chiappori and Pinquet (2003). This will be done in section 6.3.

6.2 Using an exogenous change

In this section let us study the impact on the sinistrality of the incentives change between period B and C. Let us assume in a first step that the sinistrality is stationary. As explained in the previous section, this incentive to drive carefully is stronger in period C than in period B. Let us consider the set of individuals with no bonus who reported a claim during period A. In case of absence of moral hazard, the stationarity implies that claims are equally distributed over time. In the case of the absence of moral hazard, claims are more likely to be located during period B. This situation is summarized on the following scheme:

In the case of the presence of moral hazard:

$$\Big(\overbrace{\underbrace{0-0-1}_{1st\ claim}\ \underbrace{0-1-0}_{overweighting}}^{1st\ year}\ ;\overbrace{\underbrace{0-0-0-0}_{underweighting}}^{2nd\ year}\Big)$$

In the case of the absence of moral hazard:

$$\Big(\underbrace{0-0-1}_{1st\ claim}\underbrace{0-0-0\ ;\ 0-1-0-0}_{equal\ weighting}\Big)$$

By grouping by months claims of the population defined above, it should be possible to test the presence of moral hazard. Concerning the assumptions, two points must not be forgotten. Firstly, the description above assumes that without moral hazard, the sinistrality is stationary across time. Secondly, the fact to consider insurees during two year most not blur the fact that some insurees may exit the panel, and the test of moral hazard is valid only if there is no selection effect associated to these exits. However, it is difficult to imagine a selection effect at this level.

Let us consider the population of insurees with no bonus such that the indicator to report a claim implying a third party damage or injury is equal to one exactly once during the first 6 months and equal to one exactly once during the 18 following months (in the following, positivity of the indicator and claim will be associated).

In order to have an idea of the impact of the nonstationarity of the sinistrality, let us represent the contract renewals by month.

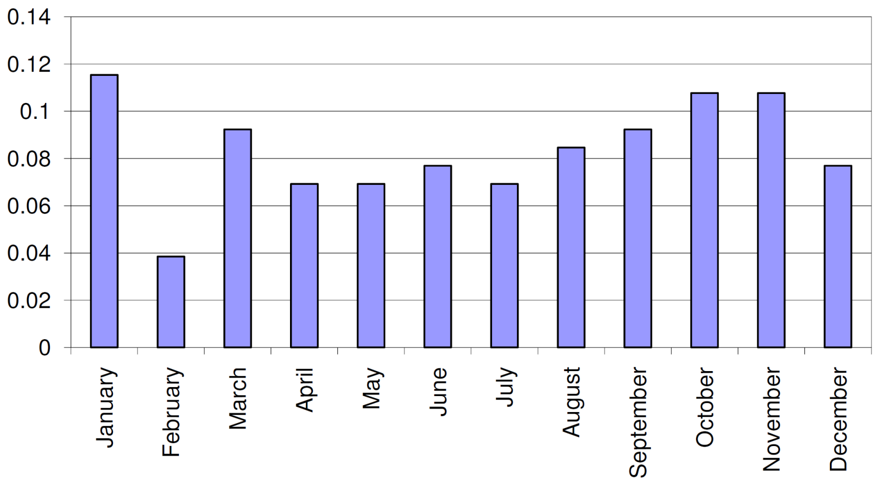

Figure 4: Repartition by month of the contract renewal date.

Renewal is quite well distributed across months, which means that by creating a cylindric panel from this population, the seasonality effect should be reduced.

Always concerning this population let us represent the repartition across time of the second claim.

Data are quite volatile due to the low number of individuals satisfying all imposed conditions, however an increasing trend seems to appear during the last six months of the first year, followed by a significant negative shock, and later by a new increasing trend.

Let us recall that only monthly and not daily information is available. In the previous graph, the contract related to the first year begins in month number 1 and ends during in number 13, and the contract related to the second year begins in month number 13. Consequently the information related to month number 13 is difficult to use as it is impossible to know whether related claims occur before or after the contract renewal date. The first month for which it is sure that all claims occur after the contract renewal date is month number 14.

Let us group claims by quarter to get less volatile quantities. The third quarter consists of months 7, 8 and 9, and the fourth quarter consists of months 10, 11 and 12. Due to the lack of information concerning

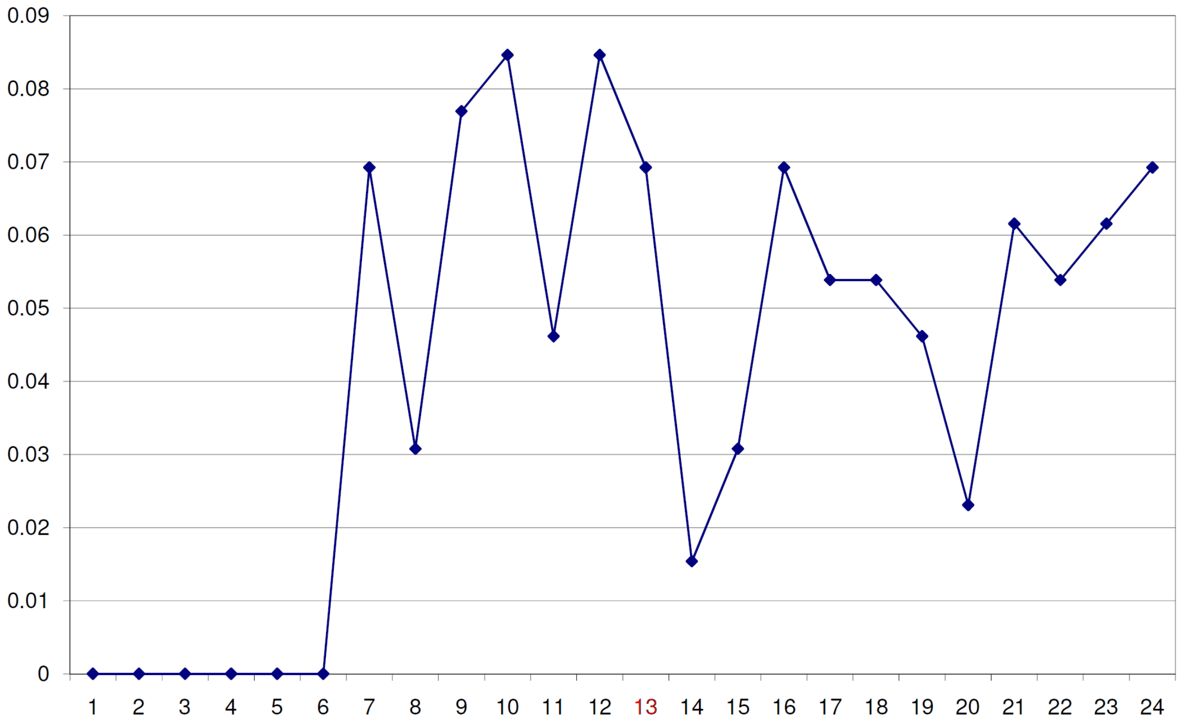

Figure 5: Repartition by month of the second claim (observed frequency in Y-axis).

month number 13, the fifth quarter consists of months 14, 15 and 16, the sixth quarter consists of months 17, 18 and 19 and the seventh quarter consists of months 20, 21 and 22. In the following graph, the X-axis represents the quarter considered and the Y-axis represents the number of claims occurred in this quarter.

We visually see a negative shock between the fourth and the fifth quarters, whereas other changes between quarters are less strong. To have a quantitative idea of the significance of the change, let us implement a proportion equality test. Let us consider only the fourth and the fifth quarters, under the corpus of assumptions mentioned above, without moral hazard, claims should be equally distributed between these two quarters. So the probability for a claim to be located in the fifth quarter should be $p = \frac{1}{2}$. Let us test this hypothesis considering that the number of claims in a quarter is equal to the realization of the sum of random variables following a Bernoulli distribution with parameter p. Considering that the number of realizations n is big enough to use the central limit theorem, we reject the proportion equality at level α if:

$$|p_{emp} - p| > u_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$$

Where $u_{\frac{\alpha}{2}}$ is the $1 - \frac{\alpha}{2}$ quantile of the standard normal distribution.

From the data we find $p_{emp} = \frac{15}{15+28} \approx 0.3488$. With $p = \frac{1}{2}$, $n = 43$ and $\alpha = 5\%$, we obtain

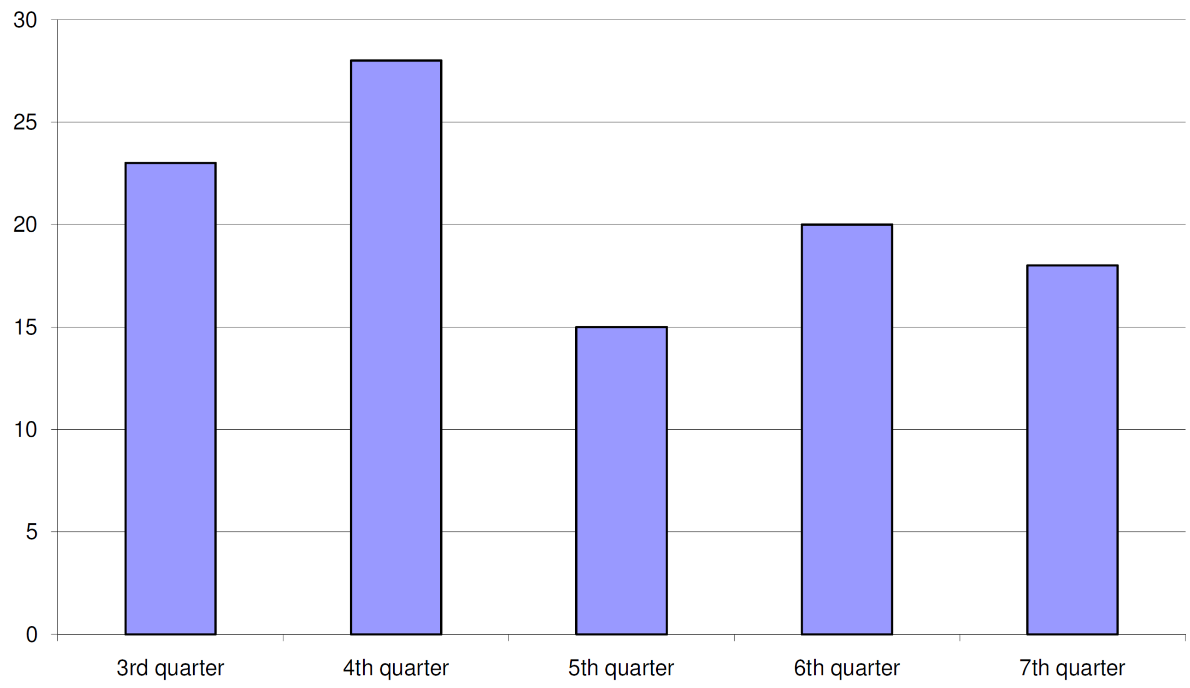

Figure 6: Repartition by quarter of the second claim (observed claim number in Y-axis).

$u_{\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} \approx 0.1494$, hence the acceptation range $[0.3506; 0.6494]$. Consequently at level $\alpha = 5\%$ we just reject the equality of the proportions. This p-value is 0.0474.

In a more natural way we can also use a chi-2 test, which is in fact a different expression of the same problem. The chi-2 distance is expressed as:

$$s = \sum_{i=1}^{k} \frac{(n_i^{emp} - n_i^{th})^2}{n_i^{th}}$$

And from the data we have:

$$s = 3.9302$$

As we have only two classes, the chi 2 distance at 95% for one degree of freedom is 3.84. Consequently we reject the equality of proportion from the chi 2 test. We could have considered directly the p-value which is again 0.0474 and conclude in this way to reject the null hypothesis at level 5%.

The non control of the stationarity limits the scope of the tests; however, if we go back to the graph representing the monthly repartition of the second claim, we can see in each period B and C separately an increasing trend, and it seems difficult to explain the negative shock between these two periods by the non stationarity. On condition that assumptions are valid, the presence of moral hazard seems to be the explanation of this shock.

6.3 Using an endogenous change

As announced in section 4.1, this section focuses on the implications of the incentives change between periods B and C, and uses the repartition across time of the first and the second claim during the first year.

This last approach, which uses the first two claims, considered as an endogenous change, is based on the article written by Abbring, Chiappori and Pinquet entitled "Moral hazard and dynamic insurance data" (2003). In their article, the authors use an intensity of claims of the form:

$$\Theta(t|\lambda, N[0,t)) = \lambda.\beta(\lambda)^{N(t-)}.\Phi(t)$$

The entire article is based on the result according which, conditionally on observable and unobservable parameters, the probability of claim decreases in case of a claim in case of moral hazard, because the marginal cost of a claim increases (results coming from the multiplicative characteristics of the malus in the French experience rating system). They thus test the presence of moral hazard by $\beta < 1$ (multiplicative decreasing in case of claim) against $\beta = 1$ (no effect on the sinistrality of past claims).

In the case of our Irish data, the experience rating system does not allow us a priori to obtain a simple conclusion concerning the marginal cost of a claim and consequently the impact of incentives. However, as the experience rating system gives a bonus and no malus, we can reuse the population described in the previous section (insurees with no bonus) and we use the fact that incentives to drive carefully decrease after a first claim. We use a model of this form:

$$\Theta(t|\lambda, N[0,t)) = \lambda.\beta(\lambda)^{1_{1er\ claim}}.\Phi(t)$$

In the case of the presence of moral hazard, we should have $\beta > 1$ (more claims after the first one because of fewer incentives to drive carefully) and in the case of the absence of moral hazard $\beta = 1$ (in the previous formula $1_{1er\ claim}$ is the indicator that a first claim occurs).

In their article, the authors consider many approaches and in addition to tests relying on stationarity assumptions, describe tests using no stationarity assumptions. One of them consists of considering two cumulative distribution functions. On the one hand there is the cumulative distribution function of the first claim among the subpopulation of insurees who report exactly one claim in the year:

$$H_1(t) = \mathbb{P}(T_1 \leq t | N(T) = 1)$$

(T_1 is the time of the first claim and $N(T)$ is the number of claims de the insuree in the year)

On the other hand there is the cumulative distribution function of the second claim among the subpopulation of insurees who report exactly two claims in the year:

$$H_2(t) = \mathbb{P}(T_2 \leq t | N(T) = 2)$$

In the case of the absence of moral hazard, they show that $H_1^2 = H_2$, whereas in the case of the presence of moral hazard $H_1^2 < H_2$ (since $\beta < 1$). H_1 and H_2 are respectively estimated by:

$$\hat{H}_{1,n} = \frac{1}{M_{1,n}} \sum_{i=1}^{n} 1_{T_{1,i} \leq t \ et \ N_i(T)=1}$$

and

$$\hat{H}_{2,n} = \frac{1}{M_{2,n}} \sum_{i=1}^{n} 1_{T_{2,i} \leq t \ et \ N_i(T)=2}$$

with $M_{k,n} = \sum_{i=1}^{n} 1_{N_i(T)=k}$

And we test the equality $H_1^2 = H_2$ from the Kolmogorov-Smirnov distance of $H_1^2 - H_2$.

We can reuse this test for the Irish data, which do not rely on stationarity assumptions.

Let us reuse the two subpopulations described above (subpopulation of insurees who reported exactly one claim in the year and those of insurees who reported exactly two).

We can preliminarily represent the monthly repartition of the renewal date in order to know if months are equally represented.

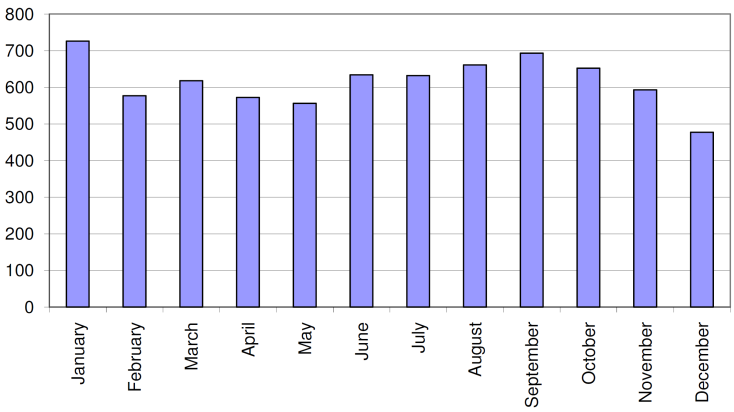

Figure 7: Repartition by month of the renewal date among the population of insurees who report exactly one claim in the year.

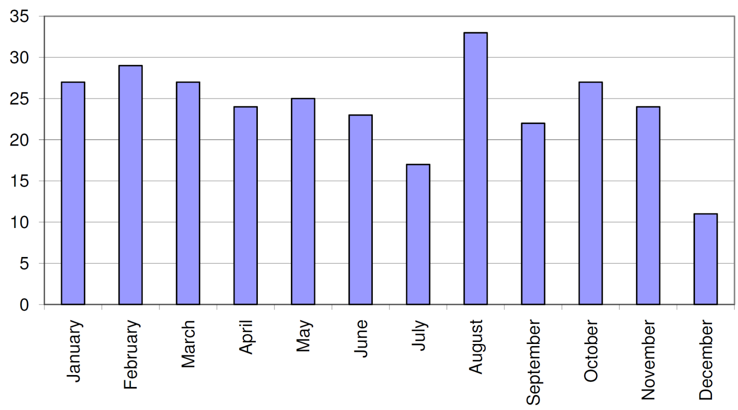

Figure 8: Repartition by month of the renewal date among the population of insurees who report exactly two claims in the year.

The repartition is quite homogeneous, more for the first subpopulation because there are more insurees in it, leading to less aggregate volatility.

Let us represent now the empirical estimations of H_1 and H_2 from the estimators mentioned above.

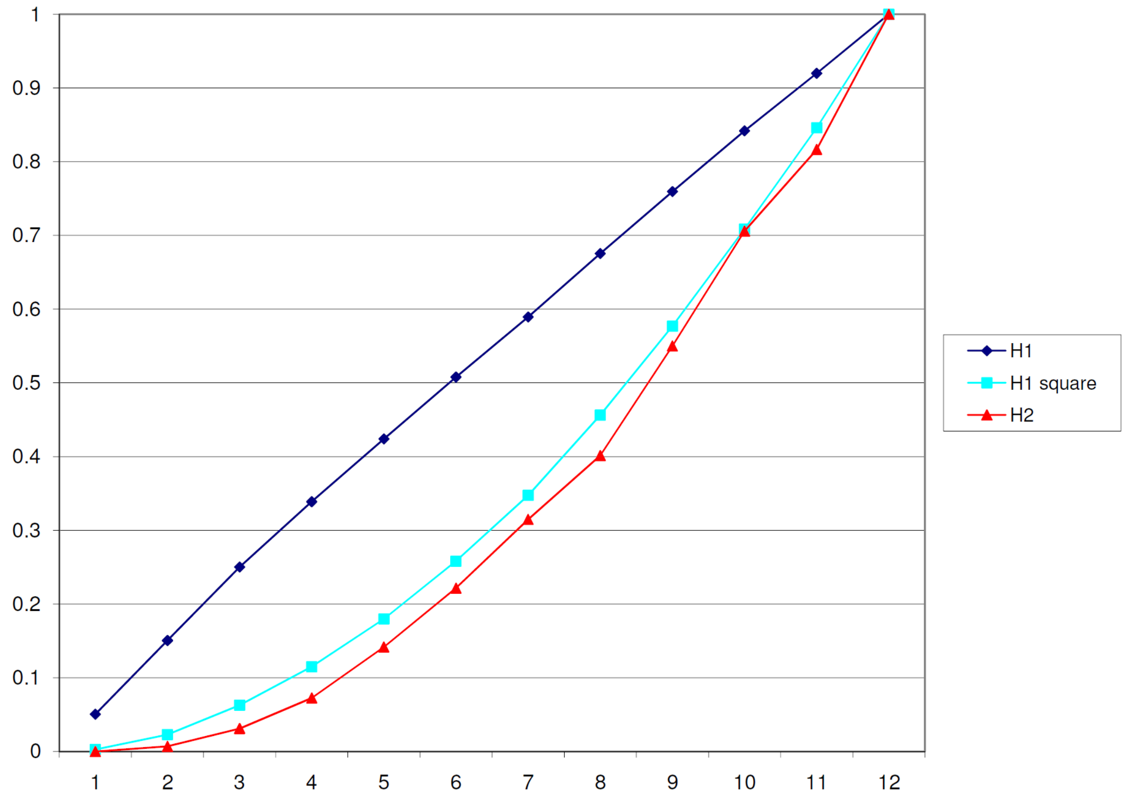

Figure 9: Empirical distribution of H_1, H_1^2 and H_2.

We visually see that $H_1^2 > H_2$, implying that $\beta > 1$ and consequently pointing toward of the presence of moral hazard.

Before testing the equality $H_1^2 = H_2$, let us assume that we have stationarity. We could directly test moral hazard by testing the equirepartition of the first claim. Let us represent the time localization of this first claim and let us recall that month number 1 corresponds to the month during which the contract begins. Consequently the effective number of days related to the first year is unknown, which is why the value related to month 1 is difficult to use (for the same reason the month number 13 was difficult to use is section 6.2). Let us test the uniformity of the time repartition on the 11 interior months with a chi 2 test.

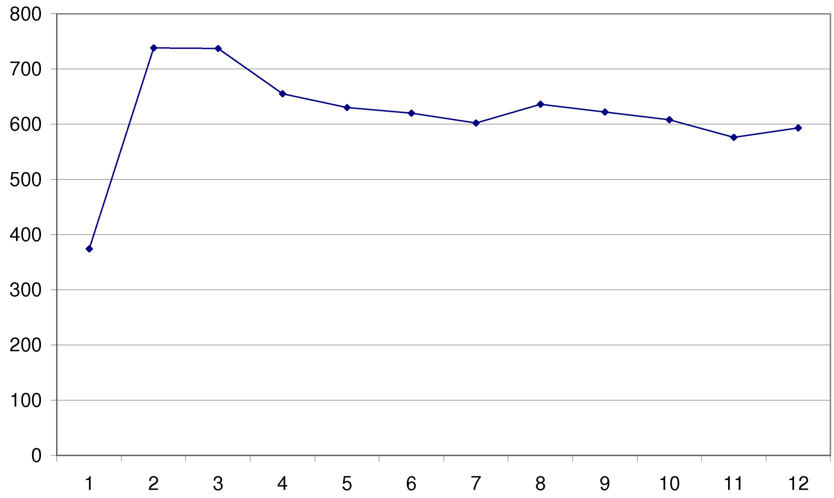

Figure 10: Repartition by month of the first claim date among the population who report exactly one claim in the year.

The chi 2 distance is 45.1521. At a 5% level, the maximum distance if the statistic follows a chi 2 distribution is 18.31. Consequently we largely reject the uniform character of the repartition. The p-value is strictly inferior to 0.0001. However, we cannot know if the rejection of the uniformity of the repartition is due to the presence of moral hazard or to non stationarity. In particular a learning phenomenon (individuals learn how to drive better their cars), leading to a decrease in the sinistrality is plausible.

Let us go back to the test of $H_1^2 = H_2$. Let us consider the Kolmogorov-Smirnov between $\hat{H}_1^2$ and $\hat{H}_2$:

$$K_n = \sup_{t \in [0,1]} |\hat{H}_{1,n}^2 - \hat{H}_{2,n}|$$

In the case of the absence of moral hazard but without any assumption on stationarity, authors show that asymptotically and conditionally on $m_1 = M_{1,n}$ and $m_2 = M_{2,n}$:

$$K_n \sim \sup_{t \in [0,1]} |\hat{U}_{1,m_1}^2 - \hat{U}_{2,m_2}|$$

where $\hat{U}_{1,m_1}$ and $\hat{U}_{2,m_2}$ are respectively cumulative distribution functions of m_1 and m_2 realisations of a uniform distribution on $[0, 1]$.

K_n does not follow an usual distribution. It is thus necessary to use simulations to be able to tabulate numerically the distribution and in fine to test the equality $H_1^2 = H_2$. In the Irish data, $m_1 = 7996$ and $m_2 = 340$. Drawing $m_1 = 7996$ and $m_2 = 340$ realisations of a uniform distribution on $[0, 1]$ allow us to have one realisation of K_n. By obtaining 1000 realisations of K_n, it is possible to get a simulated shape of its density and its cumulative distribution function.

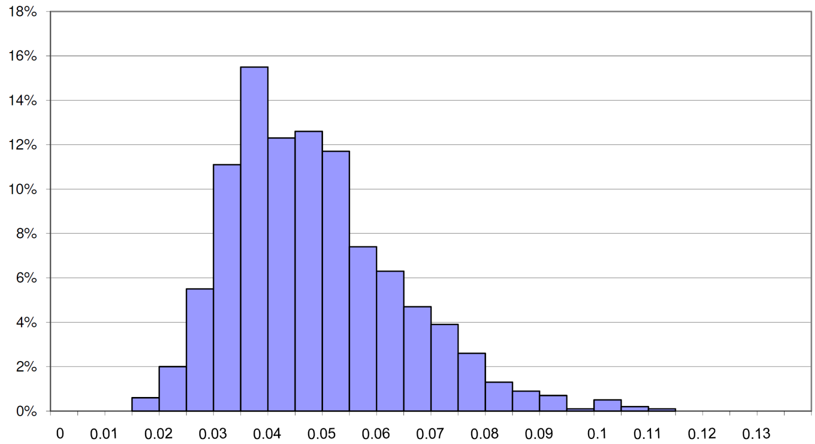

Figure 11: Histogram of the simulated distribution of K_n.

Main quantiles coming from simulations are summarized in the following table:

Quantile	Valeur
99%	0.0937
95%	0.0783
90%	0.0707
75% Q3	0.0573
50% Median	0.0459
25% Q1	0.0370
10%	0.0309
5%	0.0280
1%	0.0218

With the Irish data, we find:

$$K_n = 0.0548$$

Details of computations are given in the following table:

Month number	Number of 1st claims	Cumulated number of 1st claims	Number of 2nd claims	Cumulated number of 2nd claims	H_1^2	H_2	$\|H_1^2 - H_2\|$
1	374	374	0	0	0,0026	0,0000	0,0026
2	738	1 112	2	2	0,0226	0,0069	0,0157
3	737	1 849	7	9	0,0626	0,0311	0,0314
4	655	2 504	12	21	0,1148	0,0727	0,0421
5	630	3 134	20	41	0,1798	0,1419	0,0379
6	620	3 754	23	64	0,2580	0,2215	0,0365
7	602	4 356	27	91	0,3474	0,3149	0,0325
8	636	4 992	25	116	0,4562	0,4014	0,0548
9	622	5 614	43	159	0,5770	0,5502	0,0268
10	608	6 222	45	204	0,7087	0,7059	0,0028
11	576	6 798	32	236	0,8460	0,8166	0,0294
12	593	7 391	53	289	1,0000	1,0000	0,0000
							$K_n = 0,0548$

Consequently, at the level of 5%, we cannot reject the equality $H_1^2 = H_2$. The p-value is 0.292. This result is reported on the cumulative distribution function graph which comes from simulations of K_n:

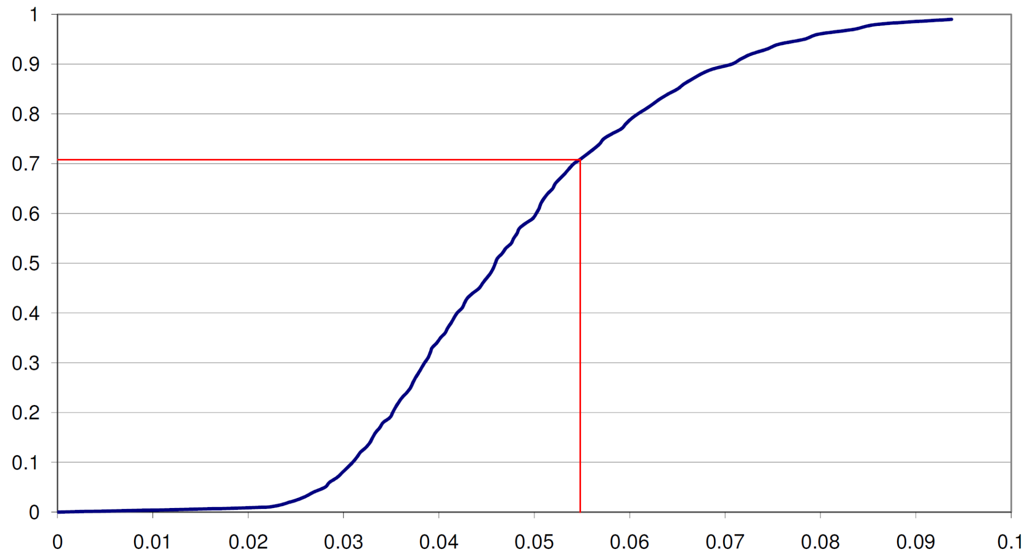

Figure 12: Simulated cumulated distribution function of K_n.

In order to be able to reject the hypothesis of equality at level 5%, we would have needed a distance greater that 0.0783. However, it is encouraging to observe that the sign of $H_1^2 - H_2$ conforms to the model prediction in the case of moral hazard.

Conclusion

The empirical puzzle of the distinction between moral hazard and adverse selection was once again confronted by data thanks to the panel provided by our major Irish insurer. After describing the problem, available data and the experience rating system in Ireland, part 5 allowed us to test the presence of information asymmetry from the partial correlation between the sinistrality and the coverage, using static data. Part 6 focused on the test of the presence of moral hazard using two approaches: the first one, simpler but more demanding in terms of assumptions, and the second one based on the article Abbring, Chiappori and Pinquet (2003), which does not require a stationarity assumption.

Results from part 5 allowed us to reject the null hypothesis of absence of information asymmetry. Section 6.2 tended to reject the null hypothesis of absence of moral hazard, but under some demanding assumptions on sinistrality. Section 6.3, with no assumption on sinistrality, led to the same conclusion as the article previously mentioned and did not allow us to reject the null hypothesis of absence of moral hazard, the Kolmogorov-Smirnov distance based on the difference between H_1^2 and H_2 being inferior to the threshold at 5%, obtained by simulations of the distribution K_n, despite the fact that the sign of $H_1^2 - H_2$ is the one predicted by the model in case of presence of moral hazard.

Possible extensions are numerous, either by taking into account more data, or by using finer tests, or by using other predictions of the presence of moral hazard, where the use of the protection state of the bonus level may be an advantage.

References

[1] Abbring J.H., P.A. Chiappori, J.J. Heckman and J. Pinquet (2003), ”Adverse Selection and Moral Hazard in Insurance: Can Dynamic Data Help to Distinguish?” *Journal of the European Economic Association*, No. 1, 512-521.

[2] Abbring J.H.,P.A. Chiappori, and J. Pinquet (2003), ”Moral Hazard and Dynamic Insurance Data” *Journal of the European Economic Association*, Vol. 1, No. 4, 767-820.

[3] Cardon J., and I. Hendel (1998), ”Asymmetric Information in Health Insurance : Evidence From the National Health Expenditure Survey”, Mimeo, Princeton University.

[4] Chiappori, P. A., F. Durand and P.Y. Geoffard (1998), ”Moral Hazard and the Demand for Physician Services : First Lessons from a French Natural Experiment”, *European Economic Review*, 42, 499-511

[5] Chiappori P. A., P.Y. Geoffard and E. Kyriadizou (1998), ”Cost of Time, Moral Hazard, and the Demand for Physician Services”, Mimeo, University of Chicago.

[6] Chiappori P.A., B. Jullien, B. Salanié and F. Salanié (2001), ”Asymmetric information in insurance: A robust characterization”, Mimeo, University of Chicago.

[7] Chiappori P.A., B. Jullien, B. Salanié and F. Salanié (2003), ”Asymmetric Information in Insurance: Some Testable Implications”.

[8] Chiappori P. A. and B. Salanié (1997), ”Empirical Contract Theory: The Case of Insurance Data”, *European Economic Review*, 41, 943-51.

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

[10] de Garidel-Thoron T. (2005), ”Welfare-Improving Asymmetric Information in Dynamic Insurance Markets”, *Journal of Political Economy*, 113, 1, 121-150.

[11] Dionne G. (2000), *Handbook of Insurance*, Kluwer Academic Publishers.

[12] Dionne G., C. Gouriéroux and C. Vanasse (1999), ”Evidence of adverse selection in automobile insurance markets”, in G. Dionne and C. Laberge-Nadeau, editors, *Automobile Insurance: Road Safety, Insurance Fraud and Regulation*, Kluwer Academic Publishers, Boston, 13-46.

[13] Dionne G., 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.

[14] Dionne G., M. Maurice, J. Pinquet and C. Vanasse (2001), "The role of memory in long-term contracting: Empirical evidence in automobile insurance", Mimeo, HEC, Montreal.

[15] Dionne G., Michaud P.-C. and Dahchour M. (2004), "Separating Moral Hazard from Adverse Selection and Learning in Automobile Insurance: Longitudinal Evidence from France", Canada Research Chair in Risk Management, Working Paper No. 04-05.

[16] Dionne G. and C. Vanasse (1992), "Automobile insurance ratemaking in the presence of asymmetrical information", *Journal of Applied Econometrics*, 7, 149-165.

[17] Eeckhoudt L., C. Gollier and H. Schlesinger (2005), *Economic and Financial Decisions under Risk*, Princeton University Press.

[18] Geoffard P.-Y., L. Gardiol and C. Grandchamp (2005) "Separating Selection and Incentive Effects in Health Insurance", CEPR Discussion Paper No. 5380.

[19] Heckman J.J. (1981), "Heterogeneity and state dependence", in S. Rosen, ed., *Studies in Labor Markets*, University of Chicago Press.

[20] Hendel I. and A. Lizzeri (1999), "The role of commitment in dynamic contracts: Evidence from life insurance", Working Paper, Princeton University.

[21] Henriet D. and C. Rochet (1986) "La logique des systèmes de bonus malus en assurance automobile : une approche théorique", *Revue Economique*, 1986, 132-152.

[22] Holly A., L. Gardiol, G. Domenighetti and B. Bisig (1998), "An econometric model of health care utilization and health insurance in Switzerland", *European Economic Review*, 42, 513-522.

[23] Holmström B. (1979), "Moral Hazard and Observability", *The Bell Journal of Economics*, Vol. 10, No. 1, 74-91.

[24] 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.

[25] Rothschild M. and J. Stiglitz (1976), "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information", *The Quarterly Journal of Economics*, Vol. 90, No. 4, 629-649.

[26] Richaudeau D. (1999), "Automobile insurance contracts and risk of accident: An Empirical Test Using French Individual Data", *The Geneva Papers on Risk and Insurance Theory*, 24, 97-114.

[27] Shapiro C. and E. Stiglitz (1984), "Equilibrium Unemployment as a Worker Discipline Device", *The American Economic Review*, Vol. 74, No. 3, 433-444.