HEDG

HEALTH, ECONOMETRICS AND DATA GROUP

THE UNIVERSITY of York

WP 14/16

"They do know what they are doing … at least most of them." Asymmetric information in the (private) disability insurance

Martin Spindler

August 2014

york.ac.uk/res/herc/hedgwp

“They do know what they are doing ... at least most of them.” Asymmetric Information in the (private) Disability Insurance*

Martin Spindler
Munich Center for the Economics of Aging (MEA), Munich, Germany
and
Massachusetts Institute of Technology, Cambridge, USA

April 15, 2014

Abstract

In this paper we analyze asymmetric information in the (private) disability insurance, which has not been analyzed before in the literature, but covers one of the most important risks faced by individuals in modern society, namely the loss of human capital. We show that there is asymmetric information, but the extent depends on the amount of coverage. Most importantly, the test with ‘unused’ observables allows us also to establish the existence of adverse selection which is of importance for contract design and public policy. Moreover, the option of choosing an annual adjustment of the insured sum, which is not used for risk classification, has strong predictive power both for the occurrence of an accident and the chosen coverage, although it should be irrelevant from the point of theory. This result shows new ways to design contracts and variable selection for risk classification. In contrast to most previous studies, we also explicitly take into consideration unobserved heterogeneity by applying finite mixture models and so called ‘unused’ observables.

Keywords: Asymmetric Information, Disability Insurance, Accident Insurance, Unused Observables, Positive Correlation, Finite Mixture Model.

JEL classification codes: D82, G22, C12, C14.

*Correspondence to: Martin Spindler, Massachusetts Institute of Technology, Cambridge USA E-mail: mspindl@mit.edu. Spindler gratefully acknowledges financial support from the Deutsche Forschungsgemeinschaft through GRK 801 and especially from The Geneva Association. I am indebted to Jochen Kneiphof and Li Zhang for providing me with the data set. I am very grateful to Bernard Salanié for his most valuable comments and Bettina Gruen for her patience and great assistance in using the R package flexmix. I thank participants of the Econometric Society North America Summer Meeting 2012 in Evanston, the annual meeting of the German Economic Association 2012 in Göttingen and the annual meeting of the German Statistical Association 2012 in Vienna for valuable comments and discussions.

1 Introduction

Asymmetric information is an important phenomenon in many markets and in particular in insurance markets. Testing for asymmetric information has become a very important issue in the literature over the last two decades, since it allows for testing theoretical predictions and of outlining new directions for research.[1] Two shortcomings are currently still present in this emerging field: One is that many branches of insurance have not been analysed yet, although recent studies show that there is no general answer as to whether there is asymmetric information in insurance markets or not. It depends on the insurance, or risk, under consideration, and on the institutional and contractual design. The second shortcoming relates to the test strategy. DeMeza and Webb (2001) show that if insurees differ in risk aversion (preferences) and if risk averse individuals are, e.g., more cautious, then selection on preferences might superpose the selection on the risk type and this might lead to a reduction in the strength of the positive correlation between risk and coverage.[2] Therefore, the results of zero correlation must be interpreted with caution.

In this paper we present the first results for the accident (disability) insurance, which—as far as we know—has never been analyzed in the literature, although this kind of insurance covers risks which are really essential. As the employed data set is new in the literature and proprietary, we give detailed descriptive statistics which might be interesting for the reader in itself.

Most importantly, the existence of unused observables offers also a possibility to disentangle adverse selection and moral hazard as shown by Finkelstein and Poterba (2006). The variable age, which is collected but not used by the insurance company, is both correlated with risk and coverage. In medical science and accident research it is well established that age and the probability of an accident are correlated and show a typical pattern. Therefore, external information shows that age is associated with the risk occurrence for other reasons than for coverage and this allows us to establish the existence of adverse selection. More details are given in the following sections.

In our analysis, we try to control for heterogeneity, e.g., differences in risk aversion, by applying finite mixture models and so called 'unused' observables, which are presented in Section 6.4 and 6.3, respectively.[3] As mentioned above, asymmetric information with respect

[1] Cohen and Siegelman (2010) give a survey of recent developments in this field.

[2] Chiappori and Salanié (2012) point out that such an 'advantageous' selection can lead (under certain conditions) only to a zero or still positive correlation in equilibrium in competitive markets but not to the negative correlation which is frequently postulated in the literature.

[3] Unused observables have been introduced in a seminal article by Finkelstein and Poterba (2006). Gan et al. (2011) use finite mixture models for testing for asymmetric information in insurance markets.

to several characteristics, e.g., risk and risk preferences, might dilute the positive correlation property. Therefore, finite mixture models are an appropriate instrument for detecting different types of individuals that are similar in their behavior. As a by byproduct, we will also see how different test procedures differ in their results when applied to real data.

Moreover, we find that the option of choosing an automatic annual adjustment of the coverage is strongly correlated with both the level of coverage and the probability of suffering an accident. It is surprising that the choice of such an option, which from a theoretical point of view should be unimportant and could also be achieved by annual renegotiation, can contain valuable information about the accident probability of an individual. This example shows that variables that, at first sight, seem irrelevant, can contain useful information which might help insurance companies mitigate the problem of asymmetric information. The option of annual adjustment, as many other variables, is not used for risk classification. Our study stresses that in the future, such irrelevant variables should also be considered for risk classification, and in the design of contracts there might also be reconsidered the selection of collected variables.

The private disability insurance which is analyzed in our paper covers three eventualities: death, invalidity and a hospital per diem. As the main reason for buying this kind of insurance is the coverage of invalidity and as this is by far the most important risk covered, we restrict our analysis to the disability component.

The paper is structured as follows: Section 2 discusses related literature and Section 3 explains in detail the institutional context of the accident / disability insurance in Germany. Section 4 presents the data set and extensive summary statistics which might of interest in itself. In Section 5 we present some insight into the testing principles and in Section 6 the econometric methods and results. Finally, Section 7 concludes.

2 Related Literature and Our Contribution

Contract theory has shown that asymmetric information can affect the efficiency of insurance markets. Therefore, testing for the existence of asymmetries in different markets has been key. But even though contract theory has a long tradition in the literature, its empirical counterpart is relatively new. Cohen and Siegelman (2010) give a recent and comprehensive survey over empirical studies in this field. One important lesson learned is that the results cannot be transferred between different markets and that the existence of asymmetric information depends on the exact risk covered.

Our first contribution is that we give a detailed analysis of the disability insurance which —as far as we know— has not yet been analyzed in the literature. The disability insurance is closely related to health insurance. Whereas health insurance covers the costs for the immediate medical treatment after an accident, disability insurance covers the long–run costs associated with severe accidents, namely disability and the associated loss of capacity to work. For asymmetric information in health insurance a detailed overview is given in Cutler and Zeckhauser (2000). Most studies covered there confirm the existence of adverse selection in health insurance. We contribute to enlarge the set of insurance branches analyzed.

Secondly, in recent studies it has been shown that unobserved heterogeneity, e.g. in risk aversion, is also important and must be controlled for. Otherwise the standard positive correlation test might deliver misleading results. Two ways to accomplish this are unused observables introduced by Finkelstein and Poterba (2006) and Finkelstein and McGarry (2006) or Finite Mixture Models as proposed by Gan et al. (2011). We will take up both approaches and compare the results. This enables us disentangling the effects of heterogeneous preferences (risk aversion type) and adverse selection (risk type).

Hendren (2013) develops a model that shows that private information, held by the potential applicant pools, might explain insurance rejection. He analyzes, amongst others, the non-group market for disability and finds a significant amount of private information held by those who would be rejected and more private information for those who would be rejected relative to those who can purchase insurance. He deduces that private information prevents the existence of large segments of this insurance market. In Germany the situation is quite different. The typical rejection of insurance coverage does not exist in the insurance market (for end-consumer[4]). The only exceptions where insurance rejections are an issue are private health insurance, where at least a basic insurance tariff has to be granted and application cannot be denied, and flood insurance, if the real estate property is located in a high-risk zone. In the private disability insurance which we analyze every applicant gets coverage according to the risk classification which comprises only two binary variables (gender and occupational class). Only in the case for exceptionally high insured sums, e.g. for artists or professional soccer player, an individual underwriting takes place but this is done by specialized insurance companies.

Accidents vs. claims. In the case of insurance contracts with deductibles, as in automobile insurance, the distinction between accidents and claims is important. Accidents which are below the deductible are usually not filed and this kind of accident is unobservable for the econometricians. One way out is to consider only accidents in the analysis which are above a certain threshold exceeding the highest deductible in any case, but this leads, technically speaking, to the comparison of truncated (conditional) distributions. In casualty

[4]In the market for insurance for companies (reinsurance) coverage might be rejected there.

insurance, all accidents are filed, so the 'accidents vs. claim problem' does not exist here, in contrast to many other types of insurance which have been analyzed recently in the literature. For the automobile insurance we refer to Spindler et al. (2013) and the reference therein.

Adverse selection and moral hazard. Most studies can only test for asymmetric information as a whole, but not for moral hazard and adverse selection separately. Especially in cross sectional data, the positive correlation can hardly be broken down into its constituent parts.[5] A way to circumvent this problem under certain circumstances is offered by the test with unused observables which is described in section 6.3 and enables us to detect adverse selection.

Moreover, we argue, that in the private disability insurance, which is analyzed in this paper, moral hazard is only of minor importance because of several reasons. This in contrast to the social disability insurance which is prone to moral hazard, as e.g. shown in Gruber and Kubik (1997).

First, it seems reasonable to assume that a disability insurance policy does not lower the diligence and prudence of the insured, as the consequences are in any case dramatic and therefore ex ante moral hazard might be of minor importance. This is in line with the approach in health economics where it is often assumed that ex ante moral hazard is negligible while the focus is on ex post moral hazard (cf. Cutler and Zeckhauser, 2000, p. 576). But in the private disability insurance ex post moral hazard is controlled by the reimbursement schedule which is quite objective.

Second, if the claim in disability insurance is sufficiently high, the claims settlement department of the insurance company requires an accident report and calls external experts to check the course of events which led to the accident and to clarify the liability and doctors to assess the size of harm according to the dismemberment schedule. Cases in which the insured event is caused on purpose or by gross negligence are extremely rare and proved cases of fraud are not contained in the data set.[6] Payments are only made if the injury / disability arises as a consequence of a documented accident. This in contrast to the social disability insurance and very important for the preconditions of moral hazard.

Third, while in the social disability insurance the number of applicants is negatively correlated with the economic situation resp. cycle, the accident probability in the private insurance is constant over time. In the period we analyze the rate of claims per insurance contracts

[5] Abbring et al. (2003) use panel data to separate moral hazard from adverse selection in the automobile insurance. They exploit the bonus malus coefficient which exists in the automobile insurance but not in other kinds of insurance.

[6] Personal communication with actuaries. More detailed figures are available, but for reasons of confidentiality, cannot be disclosed.

stays on the small level of 0.00215.

These points indicate that moral hazard is of minor importance in this market. Later we will also show that adverse selection is actually present in this market.

3 Accident / Disability Insurance

Accident insurance is essential for many insurees, as it covers risks which touch the existence of individuals, i.e., the loss of human capital. In Germany, there are two pillars of disability insurance: a compulsory disability insurance and a voluntary private disability insurance.

3.1 The Statutory Accident / Disability Insurance

The statutory disability insurance covers only risks which are related to the workplace. These are risks such as working and commuting accidents and occupational diseases. In the case of such an event, the statutory accident insurance provides both cash and non-cash benefits. The most important non-cash benefit is meeting the costs for out-patient treatment and hospital care related to the accident. In the case of accidents at work these costs are borne by the statutory accident insurance and not by the health insurance.[7] In these cases also the free choice of medical practitioner is restricted and special clinics for accident cases are provided by the responsible. The most important cash benefits are a continuation of payments and finally, if a permanent incapacity for work is diagnosed, an invalidity allowance (pension). The invalidity allowance is paid if the loss of earning capacity exceeds 20%. In the case of a full loss (100%) two third of the yearly income are paid. For smaller degrees the amount is adjustet proportionally, e.g., for a 50% invalidity only 1/3 for the yearly income is paid as disability pension. The degree of disability is determined by experts. The statutory disability insurance is financed by member companies of the employers' liability insurance association (Berufsgenossenschaft) which administrates and organizes the statutory disability insurance for different occupation classes. Everyone who is employed at these companies is covered by the disability insurance.[8] Self-employed can be insured voluntarily under special conditions.

[7] The costs for medical treatment of accidents in the leisure time are covered by the health insurance.
[8] Actually, all employed workers in Germany are covered by the statutory disability insurance.

3.2 The Private Accident / Disability Insurance

Private disability insurance usually covers disability and dismemberment, death, and a hospital per diem. These are the basic risks which are covered by default.[9] [10] In addition, some insurance companies offer additional insurance payments, e.g., reimbursements of the costs for treatment at a health resort or the costs for plastic surgery. It is important to mention that these eventualities are only covered if they arise as a consequence of an accident.

For each eventuality, i.e. disability, death and hospital stay, the insuree can choose the level of payments in the case of an occurrence of the insured event. For the hospital per diem the insuree choses the daily amount. He can also choose the amount which should be paid to the bereaved in case of his death. In the case of disability, the insuree chooses the insured sum (i.e., the coverage) for the case of a total disability (100%). For lower degrees of disabilities, which are determined according to a dismemberment schedule ('Gliedertaxe'), a proportional share of the chosen level is paid out by the insurance company. Additionally, for the case of disability, the insuree can choose between certain schedules of progression. For example the insuree can choose to obtain the full amount for total disability from disabilities of 50% onwards or to obtain twice or five times the chosen basic amount from certain levels of disability (in most cases, relatively high degrees of disability) on. The chosen progression clearly influences the insurance premium paid by the insuree.

The premium depends on these chosen levels for each eventuality. Additionally, the insurance companies can use variables such as occupation, sex, age, and so on, for risk classification and pricing. These observables also determine the insurance premium. The insurance company where our data set stems from uses only two binary variables for risk classification: gender and the riskiness of the occupation. Such a restricted risk classification is wide-spread in the accident insurance, although it might be surprising at the first glance for (insurance) economists.

Private disability insurance can be bought by both employed and self-employed and is independent from the statutory disability insurance. The underwriting process is for all applicants the same and — as mentioned above — only two variables are used for pricing, independent of the insured some or employment status. An explanation for this limited risk classification will emerge as a result of the empirical analysis.

As the most important risk covered is the disability, we will restrict our analysis to the eventuality disability.

[9] Strictly speaking, the daily payment in the case of a hospital stay consists of two parts, a hospital per diem and a convalescence allowance, but this distinction is only of minor importance for the analysis.

[10] The costs for the hospital treatment are covered either by the health insurance or the statutory disability insurance. The hospital per diem is payed for the inconvenience of the hospital stay in cash.

3.3 Comparison

As the compulsory disability insurance applies only to employees (e.g., not to the self-employed by default) and as the disability pension from the compulsory disability insurance is very limited in the case of an accident and applies only to work accidents, a private disability insurance is a very important supplement. E.g. 80% of all accidents occur outside the workplace[11] and therefore are not covered by the statutory accident insurance.

In the case of an accident, payments from the statutory and private insurance are not offset. They are paid independently, depending only on the (contractual) conditions. Therefore, for employed, the private disability insurance is additional to the statutory coverage, while for most self-employed the private disability insurance is the only source of coverage in the case of accidents.

One important difference between the statutory and private accident insurance is how the size of claim (degree of disability) is determined. In the private disability insurance the basis is the so-called dismemberment schedule and in the statutory accident insurance the impairment of the capacity to work is evaluated by experts. If an accident is associated with, e.g., the loss of a thumb, then the private disability insurance pays 20%[12] of the insured sum stated in the insurance contract. In the statutory disability it is examined how the loss of thumb impairs the employability. Therefore it is possible (and often happens) that for the same accident different degrees of capacity to work are stated depending on the occupation held. Therefore the reimbursement of the private insurance is less uncertain and provides additional security. It is a measure to hedge against the risk of the expert opinion.

In order to illustrate the institutional details introduced above we suppose an employee with average income who encounters an accident which leads to a disability of, e.g., 75% which makes it impossible to work anymore. We additionally assume that he has a private disability insurance with $200,000$ Euro insured sum and factor 2 of progression from degrees 50% onwards. If this accident happens in his leisure time the statutory insurance pays nothing. The health insurance would cover the costs for inpatient and hospital treatment. The private insurance would pay $200,000 * 0.75 * 2 = 300,000$ Euro. If the accident is related to the work place the statutory disability would pay the costs for medical treatment. Moreover, the statutory accident insurance would pay an annually disability pension. Suppose his his income was $30,000$ Euro. Than the annual pension is $30,000 * 2/3 * 0.75 = 15,000$ Euro. Additionally, he would get the payment of $300,000$ Euro of his private disability insurance.

[11] GDV (2010)

[12] This is the value of the thumb accoring to the dismemberment schedule.

4 The Data Set

4.1 Overview and Structure

For our analysis, we have had access to a proprietary data set of a German insurance company. The data set contains all contracts which were valid in the year 2005 (i.e., which were valid / under risk at January 1, 2005) or signed afterwards. These contracts are traced for a period of four years (until the end of 2008). For example, let us assume that a contract was signed before 01/01/2005 and was still active at this date. Then possibly three things can happen which are recorded in the database (not necessarily mutually exclusive):

1. One accident (or possibly several accidents) occurred during the period. Then the kind of claim and the amount of payments are recorded.

2. No claim was filed during this period.

3. The contract was terminated during this period. Then both the termination date and the claim history up to that date are recorded.

Several different persons can be insured under one contract. For example a father can take out a policy for himself, his wife, and his kids. Such constellations are also captured in the database. For each insured risk, the personal data and claim history are filed. As we want to test the risk assessment of the individual for themselves, not the ability to assess the risk of other individuals, we restrict ourselves to contracts under which only one person is insured.[13] For each insured individual, the following information is contained in the database:

1. Personal information which is partially used for pricing / rating, e.g., date of birth, sex, occupation.

2. Full information about the contract: date of signing of the contract, date of termination (if it applies), chosen amount of insurance benefit for each eventuality, annual adjustment.

3. Detailed information concerning the claims.

The data set contains approximately 2.5 million contract years. This corresponds to $n = 957,506$ insurees.

[13] Multi-person contracts are excluded and as the number of multi-person contracts is very small and as these persons (except the children) are not very different from the other insurees, this does not lead to a selection problem. The minimum age of 14.6 in Table 2 refers to single-person contract where a child is covered and the premiums are paid by the parents.

While testing for asymmetric information, the following pitfall may arise: let us assume that a contract was signed on March 1, 1990. If the contract remained valid until 2005, it is still under consideration and in the sample. But if, for example, an accident occurred in the year 2000, then the contract was removed from the data set[14] (because of administrative reasons) and is not in the sample. Therefore, there is sample selection with a tendency for 'good' risks being in the sample and 'bad risks' dropped out given the same date of signing ('attrition bias'). The whole data set would not be representative for the whole population of the insureds, as especially 'good' risks without accident are contained.

A way to circumvent this problem is to take the subsample of all contracts which were signed after January 1, 2005 and to trace them through the period. This subsample still has a size of $n = 77,125$ and will serve as "working sample". It is reasonable that the individuals might have an informational advantage concerning their accident probability but not concerning the timing of an accident.[15] Therefore, this procedure should not influence the results or distort them. All the results and analysis in Section 6 are based on this subsample. As the option of choosing an annual adjustment will become important later, we present a brief definition.

Annual adjustment. Upon signing a contract, the insurees can choose the option of annual adjustment. If they choose this option, the coverage is increased annually by a fixed percentage for each eventuality covered and the premium is adjusted correspondingly. There are two main reasons for choosing this option. One is to compensate for inflation. The other is to adapt to the standard of living. With increasing age and experience, income and living status grow and therefore higher insurance protection is preferred. This option, of annually adjusting the coverage, is chosen by 32.0% of the insurees. The average age of policy holders with this option is 35.1 years, and that without annual adjustment is 42.9. While an increase in the coverage could also be renegotiated every year when the contract is renewed, this automatic increase is especially comfortable.

Later, we will use this option as an "unused"observable. The insuree can choose to get this feature or not. This rises the question how this variable could be used for risk classification ex ante, like other demographic variables. One way to implement this would be to predefine some insured sums, where the insuree can choose from, with different insured sums available if this option is chosen (in connection with a nonlinear pricing scheme). In the current setting it could be used for an above average increase in prices in the case of an extension of contract to reflect the higher riskiness of these people as we will see later.

[14]They are classified as 'historic', a technical term from actuarial science.

[15]According to personal communication with actuaries, the accident probability rises with age, i.e., at greater ages, accidents occur more often, but this increase should concern good and bad risks comparably.

4.2 Descriptive Statistics

In this section we present detailed descriptive statistics of the data set. As this data set is new in the literature and proprietary a descriptive analysis might be of interest in itself. Although we focus on disability as the most important risk covered, we present statistics for the other eventualities.

Table 2 displays descriptive statistics of the subsample used for the analysis, i.e. all contracts signed after January 1, 2005. For completeness, descriptive statistics for the complete sample are given in the Appendix. 455 claims (i.e. accidents) were filed during the period under consideration as can be seen in Table 1. "Cost of claims" in the tables contain mainly the payments for claims in the case of an accident. A comparison of the descriptive statistics shows that the selected subsample is representative. The mean age is approximately 40 years. 57.2% of the insured are male and the dynamic adjustment is chosen by 1/3 of the insured. The highest chosen insured sum for disability is one million Euro, but the mean is about 100,000 Euro. The small mean for the cost of claims indicate that the claims distribution is highly skewed as most insurees have no accidents. This pattern is well know in actuarial science.

Table 1: Number of claims

Eventuality	Whole sample	Subsample
disability	12,901	455
death	170	11
hospital per diem	15,552	392

Figure 1 shows the distribution of age at signing the contract by gender and if annual adjustment ("dynamic option") is chosen. We see a similar pattern for men and women and that the annual adjustment is especially chosen by younger insurees. The peak at the age of 18 can be explained with the reaching the age of consent. For signing the contract at a younger age the allowance of the parents is required or the parents close the contract in the name of their child.

Figures 2 and 3 display the distribution of the chosen insured sums (higher than 5,000 Euro) and of the encountered claims, both by annual adjustment (dynamic) and gender. The claims (measured in degree of disability) are given in absolute numbers. The option dynamic is chosen by 10,620 women and 14,055 men, approximately one third of the insured. Figure 2 shows that there are some focal points in the choice of the insurance sum. Most claims are

Table 2: Descriptive statistics of the subsample

Variable	Minimum	Mean	Maximum	Standard Deviation
age	14.6	40.3	97	16.2
male	0	0.572	1	
dynamic	0	0.324	1	
insurance sum for disability	3,835	103,823	1,032,500	131,937.3
insurance sum for death	767	5,415	51,130	7,140.6
insurance sum for hospital per diem	2.5	8.3	65	9.6
cost of claims for disability	0	13.9	115,000	561.9
cost of claims for death	0	1.29	21000	126.9
cost of claims for hospital per diem	0	1.9	5,458	46.8

sample size $n = 77,125$. All figures for sums and costs are in Euro.

in the range up to 20% as Figure 3 indicates.

In order to establish the existence of adverse selection external information is needed. We will use the variable age and therefore give some insight how age and accident probability are related. Sass (2008) analyses a huge survey conducted 2004 in Germany and finds a similar pattern for both men and women: The accident probability decreases with age, but for old persons (e.g. men older than 80 years) the probability increases again. This pattern leads in whole to a negative correlation between age and risk. The microcensus for Germany (Statistisches Bundesamt, 2011) shows the same pattern for a representative sample of the whole population. Therefore there is a robust negative association between accident probability and age for the whole population (both for the insured and insured). As a side note, this pattern seems to be very stable over time and countries.[16]

To close this section, we give some facts and figures about disability insurance in Germany (GDV, 2010): in Germany, 40.8% of all households carry disability insurance. In 2009, the premium income was 6,389 million Euro and claims expenditures were 2,928 million Euro. This results in a claims ratio of 58.2%.

[16] The journal "Accident Analysis and Prevention"is a rich source for further studies which confirm this pattern.

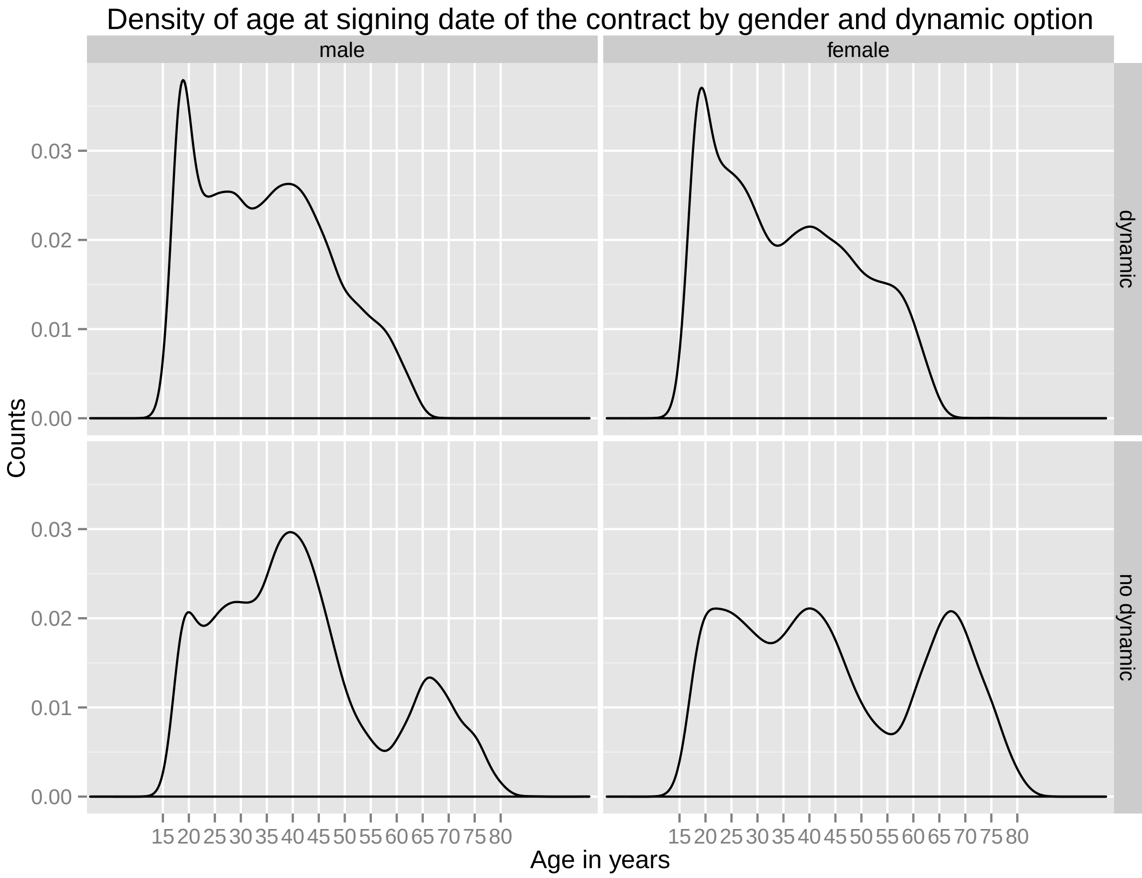

Figure 1: Age at signing the contract by gender and dynamic option.

5 The Theory of Asymmetric Information and the Basic Testing Procedures

Asymmetric information comprises two different phenomena: adverse selection and moral hazard.[17] Many equilibrium models of asymmetric information predict that insurees with more insurance coverage should be more likely to experience a loss, i.e., there is predicted a positive correlation between risk and coverage. With moral hazard, a higher insurance coverage reduces the cost of the occurrence of the insured event. This lowers the incentives for prevention or for cautious behavior, and therefore the expected loss is increased after the signing of an insurance contract.

Adverse selection means that the insured knows his risk type ex ante, i.e., before the

[17]Winter (2000) and Dionne et al. (2000) give surveys of both phenomena.

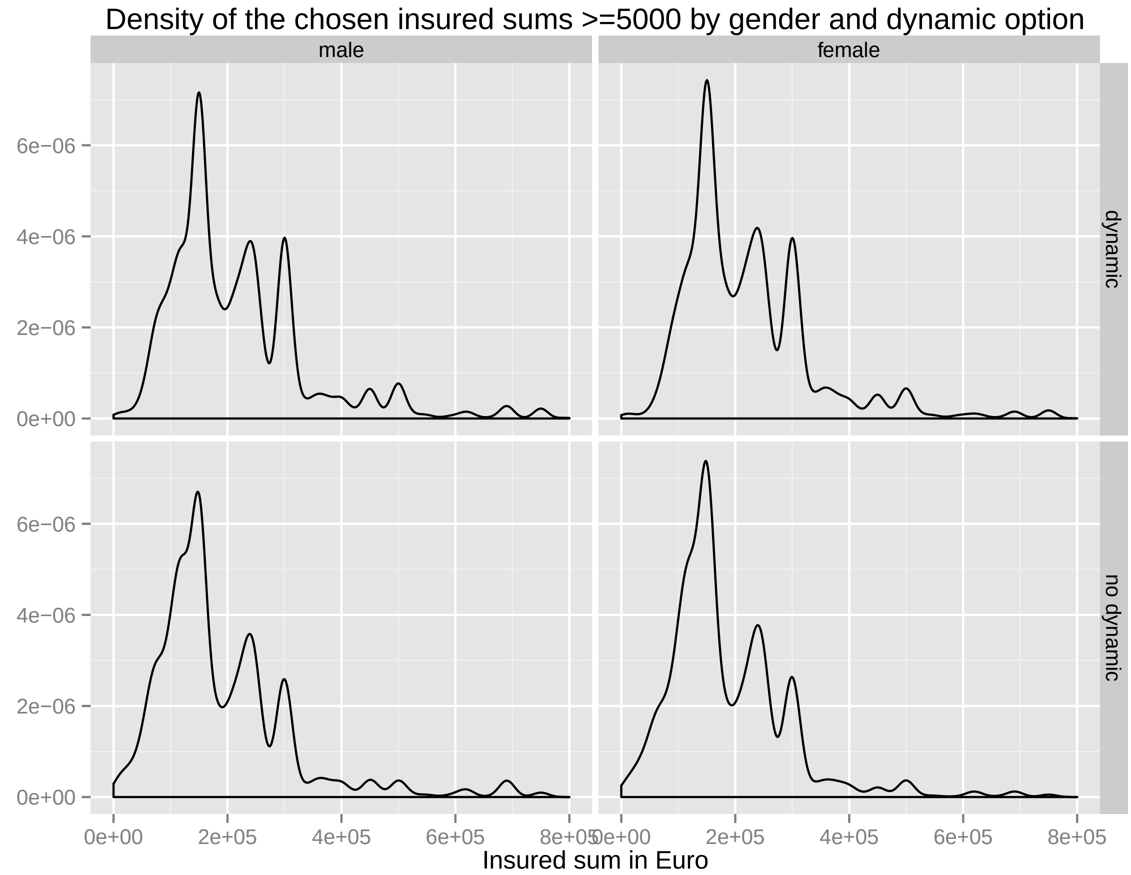

Figure 2: Chosen insured sum for disability by gender and dynamic options.

contract is signed, while the insurance company does not have this information. The insurees who know that they have a high risk will buy contracts with more coverage than the 'good' types. As 'bad' risk types have a higher marginal utility of insurance at a given price they also accept a higher per unit price for coverage and this can be exploited by the insurance companies by offering a menu of contracts to screen the different types.

Therefore the theory of asymmetric information predicts a positive correlation between risk and coverage. To identify the risk of an insuree, insurance companies use observables such as age, sex, and so on for risk classification. Thus the positive correlation property is conditional on all observables which are used for pricing.

In order to test whether there is a positive correlation, one has to set up two equations, one for the coverage (C_i) and one for the risk (or loss) (L_i). By X_i, we denote the exogenous variables which are used for risk classification. To keep the exposition simple we use linear

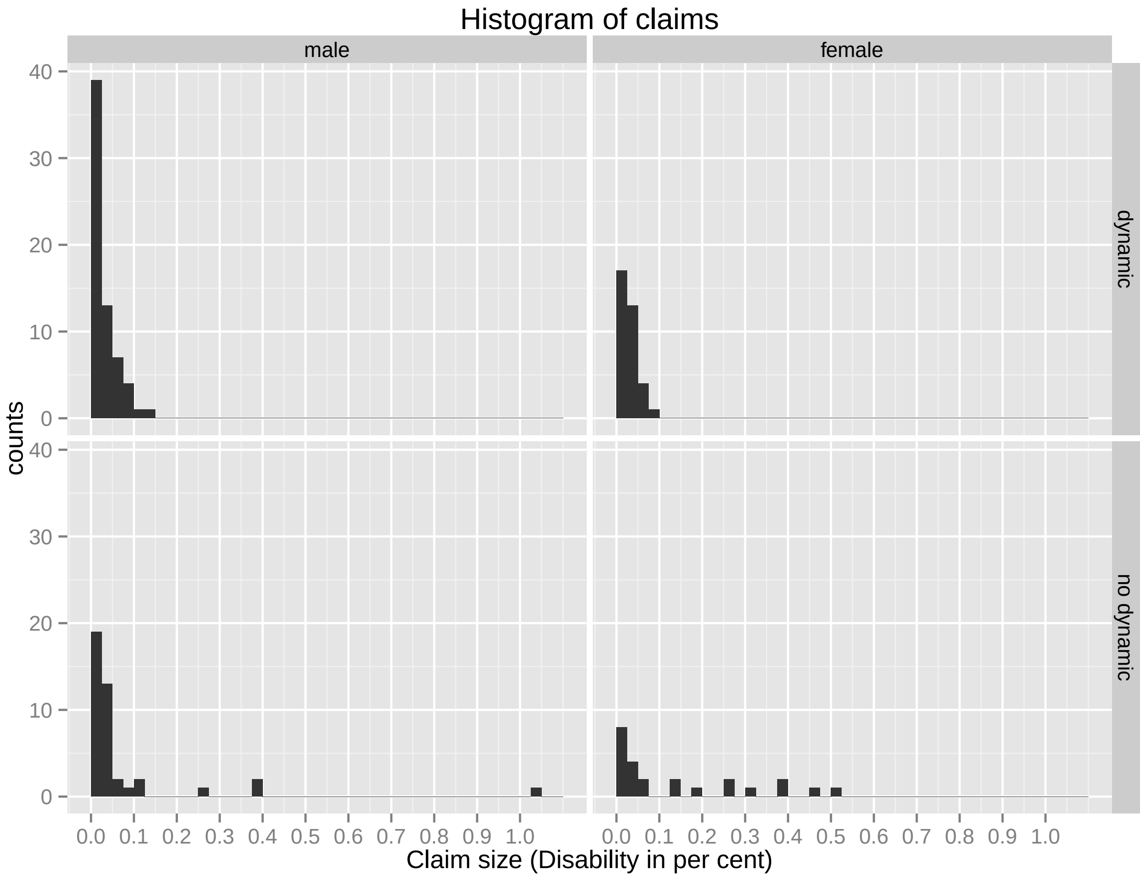

Figure 3: Absolute number of claims according to the degree of disability by gender and dynamic option.

models[18]

$$C_i = X_i * \beta + \varepsilon_i \quad (5.1)$$

$$L_i = X_i * \gamma + \eta_i \quad (5.2)$$

with error terms ε_i, η_i. Under the null hypothesis of zero correlation between risk and coverage, i.e., symmetric information, the residuals in the two equations should be uncorrelated. A significant positive correlation is an indication of asymmetric information.

This kind of test for positive correlation has been widely used in the literature: Cutler and Zeckhauser (2000) review studies in health economics, and Cohen and Siegelman (2010) give

[18] Otherwise the linear models can be regarded as approximations.

an overview of results for many different branches of insurance. The evidence for asymmetric information is not clear-cut and varies with the kind of insurance.

One important finding is that the absence of a correlation between coverage and risk can be consistent with the presence of asymmetric information. DeMeza and Webb (2001) show that when individuals have private information not only about their risk but also about their risk aversion, and when risk averse individuals are less risky (e.g., because they are more cautious), then we might also observe a zero correlation in insurance markets.[19] In other words, if individuals with stronger preferences for insurance are also of lower risk, then preference-based selection may offset risk-based selection, and the sign of the correlation is undetermined. When insurees have private information about their risk type T_i and their risk aversion A_i, both influence the unobservable error terms:[20]

$$\begin{aligned} \varepsilon_i &= T_i * \kappa_1 + A_i * \kappa_2 + \nu_i \qquad (5.3) \\ \eta_i &= T_i * \xi_1 + A_i * \xi_2 + \tau_i, \qquad (5.4) \end{aligned}$$

where ν_i and τ_i denote error terms. The principle of the correlation test is that in the case of private information about the risk type, the risk type T_i is positively correlated with coverage and risk, i.e., $\kappa_1 > 0$ and $\xi_1 > 0$. But, as stated above, if risk aversion A_i is positively correlated with coverage ($\kappa_2 > 0$) and negatively correlated with risk ($\xi_2 < 0$), then the correlation between the error terms ε_i and η_i may be zero or negative. In this case, a test for positive correlation might produce misleading results.

This shows that unobserved heterogeneity in insurance markets, especially when one tests for positive correlation, is very important and should be taken into consideration. There are several ways to test for asymmetric information and to control for heterogeneity. One is to employ unused observables, as proposed in Finkelstein and Poterba (2006), or to use finite mixture models to account for differences in risk aversion as Gan et al. (2011). We will explain these testing procedures in the following sections.

6 Methods and Results

This section gives a short review of the methods applied, customizes them to disability insurance and presents the results. First, we present shortly the basic tests for positive correlation

[19] In the literature it is claimed that advantageous selection might lead to a zero correlation, but Chiappori and Salanié (2012) argue that in competitive markets only a zero or positive correlation between risk and coverage is possible.

[20] This exposition follows Finkelstein and Poterba (2006).

and summarize the results. The detailed results are presented in the Appendix. Second, we give a summary of the test of unused observables (Finkelstein and Poterba 2006) and display the results. Thirdly, we introduce the finite mixture models which are the most general class and nest the other ones. Finally, we compare the methods and results.

We present only the results for the occurrence of disability. The danger of disability is the main reason why insurees buy disability insurance. In the case of disability insurance, only sex and the riskiness of the occupation which is given by two categories are used for pricing by the insurance company. Other variables such as age, place of residence, and so on, are not used although they are available. Therefore we conduct the standard test procedures both on only the variables used by the insurance company ('basic setting') and on an extended set of observational variables which are available in the data set. The extended configuration also has family status, occupation, age, and whether the insuree has chosen the option to increase the insured sum automatically on an annual basis (annual adjustment). The basic idea of using the extended configuration is to test if the insurees have even more knowledge than incorporated in the unused observables (and to what extend) unused variables in the data set might decrease a potential positive correlation and help to mitigate the problem of adverse selection. This is a very important point. Moreover, these additional variables also allow us to apply the 'unused observables test'.

In the following, X denotes the exogenous variables which are used for risk classification by the insurance company. C denotes the chosen contract. In disability insurance, this is the insurance sum chosen for each eventuality, for example the sum which is paid out in the case of death, the sum for a 100% disability (together with the chosen progression or factor) or the daily payment for a hospital stay. All other conditions of the contracts are identical. L measures the risk. The risk is measured as 'ex post risk', e.g., by the number of accidents, the damage payments caused by the insuree or the degree of disability stated in the case of an accident. An index i refers to a certain individual or contract, and is omitted if there will be no confusion. In Section 6.1.1, 6.1.2 and 6.3 the variable for the choice of coverage C_i is defined as a binary proxy, in Section 6.1.3 and 6.2 as a continuous variable.

6.1 Basic Testing Procedures

The theory of asymmetric information predicts a positive correlation between risk and coverage, conditional on all observables X which are used by the insurance companies for pricing. There are several econometric procedures to test this conditional 'positive correlation property'. This can be done by using two probits or a bivariate probit model and testing for a positive correlation. A detailed description is given in Appendix A. These procedures were applied to testing asymmetric information for the first time in Chiappori and Salanié (2000) and since then have become the standard procedures in this field, even when the risk or coverage variables are continuous, as it is the case, e.g., in the automobile insurance.

In order to apply these parametric procedures, one has to define binary variables for 'risk' and 'coverage'.[21] Risk is measured as ex post risk and is defined as 1 if there was at least one accident or claim, and 0 otherwise, i.e., if there was no accident during this period. As in the disability insurance only one accident can occur the definition as binary variable fits perfectly. For the 'choice of contract', we also define a binary variable, which is set to 1 if the insured sum is above a predefined cut-off value and 0 otherwise. In order to check for robustness and to analyse whether and how the asymmetric information varies with the level of the insured sum, we vary this cut-off level.[22] The contracts are identical, only the levels of insured sums can be chosen by the insuree.

The positive correlation between risk and coverage is significant for the 'low' and 'middle' cut-off levels ($10,000$ Euro, $50,000$ Euro, $100,000$ Euro, and $200,000$ Euro) for both the basic and the extended configuration. For higher thresholds, the correlations remain significant for the basic setting, while they become insignificant for the extended configuration.[23] For the low cut-offs, the correlation is quite high (between 0.5 and 0.6) for both settings. When interpreting the results, one has to keep in mind that in large samples, even small differences become (statistically) significant, while they are economically meaningless. This might apply to the small but significant correlation of the high insurance sums in the basic setting. One important finding is that additional information might reduce or eliminate the observed asymmetric information in disability insurance.[24]

A very important finding is that the results of the basic procedures (and even the procedures introduced later) hold if the unused observables are included into the analysis ("extended configuration", e.g. Tables 7 and 8) as proposed by Fang et al. (2008). As the positive correlation remains mostly, one can conlude that the applicants know more than just age, and their information is not fully revealed by the dynamic contract choice.

[21] In the literature, several different parametric methods to test for asymmetric information have been proposed. Spindler et al. (2013) show that these procedures deliver quite consistent results, so that restricting to the correlation test is no limitation.

[22] The choice of the cut-off is driven by the empirical distribution of the chosen sums, so that a natural interpretation of low, middle, and high insured sums arises. An alternative would be to choose the cut-off according to quantiles.

[23] In the extended version, we included age also as year dummies (instead of a continuous variable), but this had only a slight influence on the results.

[24] The increase of the correlation for the value of $10,000$ might be caused by functional or distributional misspecification. This issue is discussed in detail in Su and Spindler (2011).

6.2 A Linear Regression Approach

The parametric approaches described above have been successfully applied to different kind of insurance branches, even if the variables, e.g. the chosen insured sum or the claim size, are continuous, as e.g. in the automobile insurance. But also the continuous nature of the risk and choice variable can be taken into account. For the disability insurance the chosen insured sum might be explained by the risk variables X and the number of accidents N[25]:

$$C_i = X_i\beta + N_i\delta + \varepsilon_i \tag{6.1}$$

where C_i denotes now the exact chosen insured sum of the insuree in Euro. In the case of asymmetric information the coefficient for N_i is significantly positive. Dionne et al. (2001, 2006) show that – although this approach has been applied in the literature – it is vulnerable to misspecification. Therefore they propose to control for possible nonlinearities. A correction for the setting at hand is the following procedure: In a first step a probit regression of the accident variable N on the variables for risk classification is run and for each individual the predicted accident probabilities $\hat{P}$ are calculated. In a second step the regression from above is run but with the predicted probabilities as additional regressor:

$$C_i = X_i\beta + N_i\delta + \hat{P}_i\kappa + \varepsilon_i \tag{6.2}$$

If the coefficient of N_i remains positive, existence of asymmetric information is confirmed.

A modification is to replace the number of accidents (discrete variable) by the level of disability in the case of a filed claim (continuous).

Table 3: Results – Regression with and without predicted accident probability

	Model without predicted probability				Model with predicted probability			
	Estimate	Std. Error	t value	Pr(>|t|)	Estimate	Std. Error	t value	Pr(>|t|)
constant	109762.89	901.72	21.73	0.00***	−202494.11	186743.42	−1.08	0.28
occupation	−14946.64	1248.64	−11.97	0.00***	−138371.58	73823.24	−1.87	0.06*
sex	−4674.70	1153.99	−4.05	0.00***	55789.86	36178.45	1.54	0.12
number of accidents	86847.31	6070.96	14.31	0.00***	86813.99	6070.92	14.30	0.00***
predicted probability					558.42	333.96	1.67	0.0945*

***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

When we regress the chosen insured sum on the risk variables and the number of accidents, all coefficients are highly significant and therefore the ex post risk is a significant predictor for the choice of coverage, as can be seen in Table 3, columns 2 − 5. When we control for nonlinearities, i.e. take up the predicted probabilities for each insuree based on the information

[25] As the maximum number of accidents in the data set is one, this variable is in fact binary.

available to the insurance company, the number of accidents stays highly significant, while the other variables lose their influence. These results confirm that there is a positive correlation between risk and coverage which indicates the existence of asymmetric information in this market (see Table 3, columns $6-9$). Replacing the number of accidents (discrete) by the level of disability confirms the results.

6.3 Test with Unused Observables

6.3.1 Method

This test for asymmetric information was introduced by Finkelstein and Poterba (2006). The basic principle of this test is that the existence of a variable or characteristic that is known to the insuree, but unknown or not used by the insurance company (e.g., for regulatory or legal reasons), and that is (positively) correlated with both coverage and risk is an indication of asymmetric information. In the case of symmetric information, there should not exist any variable or buyer characteristic that is correlated with both insurance coverage and the risk of loss, conditional on the risk class.

This test can be formalized and implemented in the following way (using the notation from Section 2). With a potential unused observable W, we estimate the following system:[26]

$$C_i = X_i\beta + W_i\alpha + \varepsilon_i \tag{6.3}$$

$$L_i = X_i\gamma + W_i\delta + \eta_i \tag{6.4}$$

If we reject the hypothesis $\alpha = 0$ and $\delta = 0$ simultaneously for the variable W under consideration, then there is asymmetric information.

In order to implement the test, we need the same information which is needed for the positive correlation tests introduced in the previous section, i.e., information about the coverage, risk, and the exogenous variables used for pricing by the insurance company, and additionally we need the variables which are contained in the data but not used for risk classification / pricing (the so-called unused observables). When there are one or several unused variables and for each we cannot reject the null hypothesis of joint insignificance, then this does not necessarily mean that there is no asymmetric information. This can be simply due to the fact that we do not observe all relevant (unused) variables. This is a limitation to this test, and must be taken into account when interpreting the results.

A very interesting feature of this test is that it offers the possibility to detect adverse selection. Finding an unused observable that is significant in both equations is compatible with both

[26] The exposition follows Finkelstein and Poterba (2006).

adverse selection and moral hazard. When there is external information that this characteristic is correlated with risk occurrence for other reasons than insurance coverage, this is an indication of adverse selection, and it can be excluded that moral hazard is the unique source for the asymmetric information. We will apply this principle to proof that adverse selection is prevalent in this market. A very interesting feature of this test is that it offers the possibility to detect adverse selection. Finding an unused observable that is significant in both equations is compatible with both adverse selection and moral hazard. When there is external information that this characteristic is correlated with risk occurrence for other reasons than insurance coverage, this is an indication of adverse selection, and it can be excluded that moral hazard is the unique source for the asymmetric information. We will apply this principle to proof that adverse selection is prevalent in this market.

6.3.2 Results

As mentioned before, the insurance company uses only sex and the riskiness of the occupation (binary variable) for classification. In the data set there are more variables, such as age, family status, annual adjustment of the insured sum, and so on. Therefore we are in the rare situation of having additional information and of being able to conduct the test with unused observables. In the context of testing for asymmetric information, we are searching for a variable that influences simultaneously the choice of coverage and risk, each modeled as probits.

We employ the basic models introduced in Section 6.1, i.e. bivariate probits with different cut-off values, and extend the of variables used for risk classification (gender and occupational class) separately by age and annual adjustment.

Table 4 presents the results for different cut-off values in the coverage equation. There are two unused observables which have a significant influence on both variables, age and, surprisingly, the choice of an annual adjustment of the insured sum ('dynamic'). We report the parameter estimates and the corresponding t-values for both equations for the unused variable. We omit the parameter estimates of the other variables that are included as regressors, i.e., a constant, sex, and the riskiness of the occupation class, but they are available upon request. We present only the results for the additionally taken up unused observable in Table 4.

The variable annual adjustment ('dynamic') was coded in the following way: 0 if this option was chosen and 1 otherwise.

One could suppose that age has a significant influence on the choice of coverage and also on the risk. For the choice of coverage, age has a significant influence at the 1% significance level. We find that age is negatively correlated with the occurrence of an accident and significant

Table 4: Results—unused observables

Unused Variable	Cut-off	Equation	Parameter	t-value
bivariate probit				
age	50,000	risk	-0.002	-1.43
		coverage	-0.020	-68.6^{***}
	100,000	risk	-0.002	-1.79^{*}
		coverage	-0.021	-69.6^{***}
	200,000	risk	-0.002	-1.85^{*}
		coverage	-0.012	-37.5^{***}
'dynamic'	50,000	risk	-0.428	-12.5^{***}
		coverage	-3.038	-123.7^{***}
	100,000	risk	-0.443	-13.0^{***}
		coverage	-2.146	-167.5^{***}
	200,000	risk	-0.426	-12.5^{***}
		coverage	-1.281	-114.5^{***}

***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

for the medium and high insured sums. For the survival model, reported in the appendix, age is highly significant for all cut-off values. In combination with the external information about the negative relationship between age and accident probability which is stable and holds in the whole population we can conclude that adverse selection is prevalent in this market.

Surprisingly, the decision of the insuree to choose an annual adjustment contains a lot of information for the choice of risk and coverage and is highly significant in both equations and for all thresholds.

An alternative approach is to model the equation for risk by a Cox proportional hazards regression. The occurrence of an accident can be interpreted as an event / hazard and the time until occurrence can be modeled in the context of survival analysis. We present the results in Appendix B which confirm the analysis of this section. In the survival model the correlation between age and risk is even stronger, strengthening our claim.

6.4 Finite Mixture Models

6.4.1 Theoretical Background

This section describes the Finite Mixture Model (FMM) we estimate. A short introduction into FMMs and the the EM algorithm is given in Appendix C. Gan et al. (2011) introduced finite mixture models to test for asymmetric information in insurance markets.

Before we start with the definition of mixture models, we will give a motivation for why to use them. As mentioned in the introduction, it is possible that individuals have asymmetric information with regard to different dimensions, e.g., their risk and their risk aversion. If risk averse individuals are more cautious and therefore of lower risk, this might dilute the positive correlation despite the existence of asymmetric information. Finite mixture models are an elegant way of controlling for such unobserved heterogeneity, even if the different groups are not known in advance. The kind of heterogeneity does not have to be specified by the researcher. In the example above, mixture models would make it possible to detect both groups: the risk averse individuals with low accident probability and the high risk insurees, and to check whether the positive correlation property holds in each group. Therefore the finite mixture model might give more differentiated answers in situations where the one component model gives wrong or misleading answers.

The number of components for estimation is fixed, but the number of components can be chosen according to the AIC criterion comparing FMMs with different number of components. For the data set at hand, we consider a mixture of bivariate probit models. For the bivariate probits we use the parametrization proposed in Greene (2008).

We start with a two-equation latent variable model, where the first equation describes the choice of coverage and the second one the risk as in the basic models:

$$Y_1^* = X_1'\beta_1 + \varepsilon_1, Y_1 = 1 \quad \text{if} \quad Y_1^* > 0, 0 \quad \text{otherwise,} \tag{6.5}$$

$$Y_2^* = X_2'\beta_2 + \varepsilon_2, Y_2 = 1 \quad \text{if} \quad Y_2^* > 0, 0 \quad \text{otherwise,} \tag{6.6}$$

$$\mathbf{E}[\varepsilon_1|X_1, X_2] = \mathbf{E}[\varepsilon_2|X_1, X_2] = 0 \tag{6.7}$$

$$\mathbf{Var}[\varepsilon_1|X_1, X_2] = \mathbf{Var}[\varepsilon_2|X_1, X_2] = 1 \tag{6.8}$$

$$\mathbf{Cov}[\varepsilon_1, \varepsilon_2|X_1, X_2] = \rho. \tag{6.9}$$

To construct the likelihood, we define $q_{i1} = 2Y_{i1} - 1$ and $q_{i2} = 2Y_{i2} - 1$. $q_{ij} = 1$, if $Y_{ij} = 1$ and -1, if $y_{ij} = 0$ for $j \in \{1, 2\}$. Now let

$$z_{ij} = X_{ij}'\beta_j \quad \text{and} \quad w_{ij} = q_{ij}z_{ij}, \quad j = 1, 2, \tag{6.10}$$

and

$$\rho_i^* = q_{i1}q_{i2}\rho. \tag{6.11}$$

Denoting with ϕ_2 and Φ_2 the bivariate normal density function and cumulative density function, repectively, the probability for an observation (y_{i1}, y_{i2}) is given by

$$Prob(Y_1 = y_{i1}, Y_2 = y_{i2}|x_1, x_2) = \Phi(w_{i1}, w_{i2}, \rho^*) \tag{6.12}$$

In the case of a finite mixture model with two components the probability for an observation (y_{i1}, y_{i2}) is given by

$$Prob(Y_1 = y_{i1}, Y_2 = y_{i2}|X_1, X_2) = p\Phi(w^1_{i1}, w^1_{i2}, \rho^{*,1}) + (1-p)\Phi(w^2_{i1}, w^2_{i2}, \rho^{*,2}) \tag{6.13}$$

where p denotes the mixture probability and the superscript the corresponding component. The components can vary in both the parameter vectors entering the index and the correlation ρ.

Finally we can form the log likelihood of an iid sample and maximize this function with the EM algorithm.

We close this section with a remark about identification. Although it is not possible to nonparametrically identify both prefences and risk heterogeneity, the parametric FMM with constant mixture probabilities which we employ is clearly identified.[27] Finite mixture models with constant mixing probabilities contain the single component model as a special case and allow more flexibility in any case. Therefore the basic model with and without unused observables is nested in the FMM framework.

6.4.2 Results

We implemented the mixture model in R using the package flexmix.[28] As the EM algorithm is only capable of finding local maxima, we used different starting values (random assignment, 20 trials). We repeated the procedure for different cut-off values to check its robustness and to compare the results with the results of a single component.

We present the corresponding parameter estimates stacked in a column in Table 5. Sex is coded as 0 for male and 1 for female. The occupation dummy is set to 1 for dangerous jobs and 0 for normal jobs. We conducted the FMM also for the basic configuration of the insurance company. For lack of space and as the results concerning the correlation, variables and components are comparable, we present and interpret only the results for the extended versions as it offers more detailed insights into the nature of the different types.

The last row presents the AIC:[29] for all three models it is lower than the corresponding

[27] Identification in finite mixture models has been addressed recently by Gan et al. (2011) and especially by Henry et al. (2013).

[28] The package flexmix is described in Leisch (2004) and Gruen and Leisch (2007).

[29] The BIC gives comparable results and therefore are omitted for brevity.

value for the one component bivariate probit model. Therefore, the two component model is preferable to a one component model from the point of view of model selection. We omit the results for three and four component mixture models as according to the AIC the two component model is selected. The AIC can be used to select the number of components for a fixed cut-off value, but it has only limited informative power in comparing models with different cut-off values as our primary goal is not to find the best fit but to analyse how different chosen insured sums are related to risk.

For the models with thresholds $50,000$ and $100,000$ Euro, we observe quite similar patterns: We can identify a group which has a lower risk of having an accident (Component 1), while Component 2 has a higher risk of having an accident as is indicated by the higher constant. But the choice of coverage is similar for both groups. The constants are not significantly different. Nevertheless we observe a high correlation between risk and coverage within each component. For the model with the highest threshold, the interpretation is slightly different: we observe again one group which is less risky (Component 1) than the other. The choice of contract is similar for both groups. The only difference is that for Component 2 we have a zero correlation, while for Component 1 the correlation is significantly positive.

6.5 Comparison and Remarks

The Finite Mixture Model is the most general model and nests both the corresponding basic (single component) model and even the models including "unused"observables. We find that a FMM with two components is appropriate and therefore the widely used positive correlation test is misspecified. An important question is what can be learned in the case of misspecification. If there is a positive correlation for both components then it is expected that a positive correlation is also observed in the single component model. Therefore if the single component model indicates asymmetric information this is strong evidence. But for higher insured sums we observe that in one component the positive correlations disappears and what happens in aggregation might be unclear. Similarly, if there is an unused variable which is correlated with both risk and coverage in the single component model this is an indication for asymmetric information. The other case, i.e., that an unused observable is not significant is harder to interpret and might be because of misspecification. Nevertheless, unused observables can be easily incorporated in the FMM framwork.

The second point we would like to discuss is the nature of the variable, i.e., continuous vs. discrete variables. In our case both coverage (the chosen insured sum) and risk (degree of disability) are continuous. A restriction to binary variables might lead to a loss of information, especially for the insured sum. In case of an accident, the degree is only second-rate, and

Table 5: Results—two components

	cut-off 50,000		cut-off 100,000		cut-off 200,000	
	Component 1	Component 2	Component 1	Component 2	Component 1	Component 2
risk						
constant	-3.097^{***}	-1.847^{**}	-3.040^{***}	-1.795^{***}	-3.042^{***}	-1.863^{***}
	(0.712)	(0.846)	(0.467)	(0.340)	(0.584)	(0.390)
age	0.009^{*}	-0.015^{*}	0.009^{**}	-0.018^{**}	0.008^{*}	-0.015
	(0.005)	(0.008)	(0.004)	(0.009)	(0.005)	(0.009)
sex	0.259	-0.544^{**}	0.203	-0.524^{***}	0.255	-0.520^{**}
	(0.247)	(0.258)	(0.194)	(0.200)	(0.277)	(0.236)
occupation	0.139	0.071	0.111	0.093	0.222	0.021
	(0.175)	(0.095)	(0.129)	(0.087)	(0.170)	(0.089)
coverage						
constant	0.725^{***}	0.730^{***}	0.635^{***}	0.672^{***}	-0.169	-0.213^{*}
	(0.052)	(0.069)	(0.062)	(0.077)	(0.152)	(0.114)
age	-0.020^{***}	-0.020^{***}	-0.021^{***}	-0.021^{***}	-0.013^{***}	-0.012^{***}
	(0.0004)	(0.005)	(0.0003)	(0.0007)	(0.001)	(0.001)
sex	0.098	0.082	0.077	0.078	-0.097	-0.068
	(0.093)	(0.084)	(0.077)	(0.077)	(0.137)	(0.131)
occupation	0.015	-0.017	-0.074^{***}	-0.121^{***}	-0.323^{***}	-0.263^{***}
	(0.019)	(0.065)	(0.024)	(0.046)	(0.084)	(0.052)
correlation	0.582^{***}	0.554^{***}	0.508^{***}	0.349	0.325^{***}	0.040
	(0.330)	(0.325)	(0.133)	(0.172)	(0.059)	(0.122)
AIC	$106,758.1$		$105.591,7$		$83,508.1$	

***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

the indicator variable if an accident occurred is a reasonable proxy. To take the continuous nature into account we estimate a regression model with continuous choice variable in Section 6.2 and confirm the existence of asymmetric information. A related approach is

$$C_i = X_i\beta + \varepsilon_i \tag{6.14}$$

$$L_i = \mathbf{1}(X_i\gamma + \eta_i > 0) \tag{6.15}$$

where $\mathbf{1}(\cdot)$ denotes the indicator function.

The advantage of even transforming the insured sum to a binary variable with different cut-off values is that it gives a "finer resolution" how the extent of asymmetric information varies with the insured sum. E.g. the regression approach shows that there is asymmetric information in the data and the approach with different cut-off values shows that this is especially the case for the low insured sums while for the really high insured sums which might suggest superior knowledge of the insuree do not show a positive correlation.

Therefore, in this paper we apply all approaches to give an as detailed picture as possible and see how the results vary depending on the method applied.

7 Conclusion

In this paper we analyzed the extent of asymmetric information in disability insurance, especially taking into account unobserved heterogeneity. While disability insurance covers various eventualities, we concentrated on disability, as it is the main reason for choosing this kind of insurance and covers risks which are essential for the insured. Besides the standard tests for asymmetric information, we also applied the test with unused observables and used a finite mixture model with two components. As only relatively few variables for risk classification are used in disability insurance, we are in the rare situation of being able to find variables in the data set which are not used by the insurance companies. The unused observables age and, surprisingly, the choice of an annual adjustment of the insured sum have an influence on both the choice of the insured sum and the probability of having an accident. Therefore they indicate the existence of asymmetric information even in the presence of differences in preferences (risk aversion). Moreover, external information concerning accidents related to age allows us to conclude that adverse selection is existent in this market and more important than moral hazard. This is a very important fact when deriving conclusions for contract design and public policy.

The finite mixture model indicates the existence of two different types of insurees. The concrete interpretation depends on the definition of the cut-off value used for the proxy for risk, but it is not straightforward in the sense of high and low risk averse types as in Gan et al. (2011). For each component, we found a significant positive correlation between risk and coverage, and therefore we can confirm the result of the standard model especially for low and middle range insured sums. For high insured sums, the correlation within each group is lower, and for one component we found a zero correlation between risk and coverage, therefore an absence of asymmetric information. This result is quite interesting as it delivers a deeper insight than the one component model and shows that two different types can be identified.

The fact that the variable annual adjustment, which should be irrelevant from a theoretical point of view, is correlated with the occurrence of accidents, shows that seemingly useless variables can contain valuable information which might mitigate the problem of asymmetric information in insurance markets and should be considered for designing contracts and for risk classification. This is probably the most important result of our study as it offers new ways for risk classification, pricing, and contract design.

We hope that the different tests we apply in this paper give a comprehensive picture of the extent of asymmetric information and adverse selection in this market. In future research, we would like to analyze more deeply the joint distribution and dependence structure of risk and coverage, e.g. with copulas, and to refine the finite mixture approach.

A Testing for a Positive Correlation

A.1 Methods

A.1.1 Two Probits

One approach is to define two probit models, one for the choice of the coverage C_i (either a low or a high insured sum) and the other for the occurrence of an accident L_i (either no accident or damage occurred, or at least one):

$$\begin{cases} C_i = \mathbf{1}(X_i\beta + \varepsilon_i > 0) \\ L_i = \mathbf{1}(X_i\gamma + \eta_i > 0), \end{cases} \tag{A.1}$$

where ε_i and η_i are independent standard normal errors, and β and γ are coefficient vectors (as columns). C_i is defined to be 1 if the insured sum is above a cut-off value, and 0 otherwise. L_i is 1 if at least one accident happened during the period under consideration, and 0 otherwise. The row vector X_i denotes the covariates of individual i. First, these two probit models are estimated independently and then the generalized residuals $\hat{\varepsilon}_i$ and $\hat{\eta}_i$[30] are calculated. These are required for the following test statistic:

$$W_n = \frac{\left(\sum_{i=1}^{n} \hat{\varepsilon}_i \hat{\eta}_i\right)^2}{\sum_{i=1}^{n} \hat{\varepsilon}_i^2 \hat{\eta}_i^2}. \tag{A.2}$$

Under the null of conditional independence, $\mathrm{cov}(\varepsilon_i, \eta_i) = 0$ and W_n is distributed asymptotically as $\chi^2(1)$, as shown by Gourieroux et al. (1987).
Chiappori and Salanié (1997, 2000) introduced this approach. One drawback is that information is lost as Y and Z have to be defined as binary variables. This is not problematic for the risk variable as more than one accident is extremely rare. But for the coverage C_i, one must choose a cut-off point to define the low and high insured sums. In order to check the robustness of our results, we will choose several cut-off points to see how the definition of the binary variable C_i influences the results.

A.2 Bivariate Probit

A related approach is to estimate a bivariate probit model in which ε_i and η_i are distributed as bivariate normal with correlation coefficient ρ, which has to be estimated, and then to test whether $\rho = 0$ or not. In order to test this hypothesis the Wald, Score, or LR test can be

[30] For example, the generalized residual $\hat{\varepsilon}_i$ estimates $E\left(\varepsilon_i | L_i\right)$. See Gourieroux et al. (1987) for the definition of generalized residuals in limited dependent models and applications.

used.

A.2.1 Results

In order to apply these parametric procedures, one has to define binary variables for 'risk' and 'coverage'.[31] Risk is measured as ex post risk and is defined as 1 if there was at least one accident or claim, and 0 otherwise, i.e., if there was no accident during this period. As in the disability insurance only one accident can occur the definition as binary variable fits perfectly. For the 'choice of contract', we also define a binary variable, which is set to 1 if the insured sum is above a predefined cut-off value and 0 otherwise. In order to check for robustness and to analyse whether and how the asymmetric information varies with the level of the insured sum, we vary this cut-off level.[32] The contracts are identical, only the levels of insured sums can be chosen by the insuree. In Tables 6 and 7 below, the results for the basic and extended setting are given according to the different cut-offs for the proxy of the insured sum. The test statistics in our paper are not weighted according the duration of the contract. Weighing the contracts leads to the same results and conclusions.

The positive correlation between risk and coverage is significant for the 'low' and 'middle' cut-off levels ($10,000$ Euro, $50,000$ Euro, $100,000$ Euro, and $200,000$ Euro) for both the basic and the extended configuration. For higher thresholds, the correlations remain significant for the basic setting, while they become insignificant for the extended configuration.[33] For the low cut-offs, the correlation is quite high (between 0.5 and 0.6) for both settings. When interpreting the results, one has to keep in mind that in large samples, even small differences become (statistically) significant, while they are economically meaningless. This might apply to the small but significant correlation of the high insurance sums in the basic setting.

We applied the tests relying on two probits, on a bivariate probit and a nonparametric test introduced in Chiappori and Salanié (2000) for the basic setting (CS np test). As all variables used for risk classification are binary in the basic configuration, this test is a natural choice. All these tests gave similar results as can be seen in Table 6. Although the CS np test already

[31] In the literature, several different parametric methods to test for asymmetric information have been proposed. Spindler et al. (2011) show that these procedures deliver quite consistent results, so that restricting to the correlation test is no limitation.

[32] The choice of the cut-off is driven by the empirical distribution of the chosen sums, so that a natural interpretation of low, middle, and high insured sums arises. An alternative would be to choose the cut-off according to quantiles. Although for choice of coverage a binary variable is only an approximation, this procedure is in line with the literature. Moreover, varying the cut off value should give a comprehensive insight into the nature of the relationship between risk and coverage. But as the chosen thresholds cover the set of insured sums with a quite close mesh, this approach is preferred.

[33] In the extended version, we included age also as year dummies (instead of a continuous variable), but this had only a slight influence on the results.

confirms the existence of asymmetric information we nevertheless additionally apply tests which explicitly control for differences in risk aversion and which are capable of dealing with more than only binary variables. This enables us to determine different types of heterogeneous preferences and to evaluate the effects of including not used additional variables.

One important finding is that additional information might reduce or eliminate the observed asymmetric information in disability insurance as a comparison of Table 6 and Table 7 shows.[34]

Table 6: Results—basic configuration

	cut-off point in Euro	$10,000$	$50,000$	$100,000$	$200,000$	$300,000$	$400,000$
% of contracts below		51.4	52.1	56.6	78.8	94.6	96.9
bivariate	ρ	0.5614***	0.5152***	0.3903***	0.1805***	0.1211***	0.0862**
probit	s.e.	0.0245	0.0229	0.0207	0.0213	0.0301	0.0372
	t-statistic	22.9	22.45	18.87	8.48	4.02	2.320
	LR-Test	516.0	470.1	325.3	69.0	15.4	5.1
two probits	W	298.20	275.15	190.68	35.18	7.70	2.98
CS np test	χ^2	425.8	404.2	323.3	120.9	129.0	107.5

***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

B Survial Model for the Test with Unused Observables

In our data set, the time of the occurrence of accidents is filed and therefore this can also be interpreted as survival data. Introductions to survival analysis are, amongst others, Klein and Moeschberger (2003) and Kalbfleisch and Prentice (2002).
The Cox proportional hazards regression models the hazard rate as a function of time t:

$$log\,(h_i(t)) = \alpha(t) + \beta_1 * x_{i1} + \cdots + \beta_k * x_{ik}$$

or, equivalently,

$$h_i(t) = h_0(t) + \exp(\beta_1 * x_{i1} + \cdots + \beta_k * x_{ik}),$$

[34]The increase of the correlation for the value of $10,000$ might be caused by functional or distributional misspecification. This issue is discussed in detail in Su and Spindler (2011).

Table 7: Results—extended configuration (sex, occupation, family status, age, dynamic)

	cut-off point in Euro	$10,000$	$50,000$	$100,000$	$200,000$	$300,000$	$400,000$
% of contracts below	51.4	52.1	56.6	78.8	94.6	96.9	
bivariate	ρ	0.6209***	0.5296***	0.2918***	0.0878***	0.0455	0.0108
probit	s.e.	0.0281	0.0280	0.0271	0.0234	0.0314	0.0383
	t-statistic	22.07	18.91	10.75	3.76	1.45	0.281
	LR-Test	369.4	303.0	142.0	15.31	2.122	0.080
two probits	W	286.0	262.2	176.6	30.5	6.2	2.18

***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

where $x_{i1}, \ldots, x_{ik}$ denote the covariates of individual i. The model is semiparametric because while the baseline hazard function $h_0(t)$ can take any form, the covariates enter the model in a linear way. The hazard function, which is time dependent in the Cox model, assesses the instantaneous risk of demise at time t, conditional on survival to that time. In the context of testing for asymmetric information, we are searching for a variable that influences simultaneously the choice of coverage and the hazard function, i.e., the survival function.

Table 8: Results—unused observables

Unused Variable	Cut-off	Equation	Parameter	t-value
bivariate probit				
survival model				
age	$50,000$	risk	-0.014	-4.73***
		coverage	-0.020	-68.6***
	$100,000$	risk	-0.014	-4.73***
		coverage	-0.21	-69.6***
	$200,000$	risk	-0.014	-4.73***
		coverage	-0.012	-37.4***
'dynamic'	$50,000$	risk	-0.543	-5.47***
		coverage	-3.041	-123.5***
	$100,000$	risk	-0.543	-5.47***
		coverage	-2.148	-167.5***
	$200,000$	risk	-0.542	-5.47***
		coverage	-1.281	-114.5***

***, **, * denote statistical significance at the 1%, 5%, and 10% level, respectively.

The parameters for dynamic in the survival model are interpreted in the following way: holding the other covariates constant, not choosing the annual adjustment reduces the hazard rate[35] of having an accident by a factor of $e^{-0.543} \approx 0.60$ on average—that is, by approximately 40%.

The results in Table 8 confirm the results of the analysis with two probits. We identify both age and annual adjustment as unused observables which are both highly correlated with risk and coverage.

C Finite Mixture Models

This section briefly introduces finite mixture models, which are well established in the statistical literature, especially in combination with the EM algorithm for estimation. For a more detailed introduction, refer to McLachlan and Peel (2000) or Fruehwirth-Schnatter (2006). Leisch (2004) and Gruen and Leisch (2007) also give an introduction to finite mixture models and provide an implementation in the R package flexmix.

A finite mixture model with K components is given by

$$h(y|x, w, \psi) = \sum_{k=1}^{K} \pi_k(w, \alpha) f_k(y|x, \theta_k),$$

where $\psi = (\alpha, \theta_1, \ldots, \theta_K)$ is the vector of all parameters for the mixture density $h()$. It consists of the parameter for the mixture probability α and the parameters for the separate component distributions θ_i, $i = 1, \ldots, K$. f_k denotes the density of the kth component. y denotes the response, x the predictor, and w the concomitant variables.

For the component weights π_k it is required that for all w

$$\sum_{k=1}^{K} \pi_k(w, \alpha) = 1 \quad \text{and} \quad \pi_k(w, \alpha) > 0 \quad \forall k.$$

In many applications, the mixture distributions π_k are independent of other variables and thus constant.

The most common method for maximum likelihood (ML) estimation of finite mixture models with a known number of components is the Expectation-Maximization (EM) algorithm. The EM algorithm was originally introduced for the ML estimation of incomplete data (Dempster

[35]For a definition of the hazard rate we refer to Kalbfleisch and Prentice (2002) and Klein and Moeschberger (2003). Roughly speaking, the hazard function is defined as the event rate at time t conditional on survival until time t or later (that is, $T \geq t$).

et al. 1977). The ML estimation of finite mixtures models can be interpreted in this way: each observation belongs to a certain class, for which we do not observe the component membership and therefore we are in the case of incomplete data. Formally, we define a latent variable $z_n \in \{0,1\}^K$ for each observation n which indicates the class membership, i.e., z_{nk} (the kth component of the vector z_n) is equal to 1 if the observation belongs to class k, and 0 otherwise. In the EM algorithm, these unobserved class memberships z_{nk} are treated as missing values and replaced by estimations of the a posteriori probabilities $\hat{p}_{nk}$. For a sample of N observations the EM algorithm is given by

E-step: Given the parameter estimates $\psi^{(i)}$ from the ith iteration, replace the missing data z_{nk} by the estimated a posteriori probabilities

$$\hat{p}_{nk} = \frac{\pi_k(w_n, \alpha^{(i)}) f_k(y_n|x_n, \theta_k^{(i)})}{\sum_{u=1}^{K} \pi_u(w_n, \alpha^{(i)}) f_u(y_n|x_n, \theta_u^{(i)})}.$$

M-step: Given the estimates for the a posteriori probabilities $\hat{p}_{nk}$ from the previous step, obtain the new estimates $\psi^{(i+1)}$ by maximizing

$$Q(\psi^{(i+1)}|\psi^{(i)}) = Q_1(\theta^{(i+1)}|\psi^{(i)}) + Q_2(\alpha^{(i+1)}|\psi^{(i)})$$

with

$$Q_1(\theta^{(i+1)}|\psi^{(i)}) = \sum_{n=1}^{N}\sum_{k=1}^{K} \hat{p}_{nk} \log(f_k(y_n|x_n, \theta_k^{(i+1)}))$$

and

$$Q_2(\alpha^{(i+1)}|\psi^{(i)}) = \sum_{n=1}^{N}\sum_{k=1}^{K} \hat{p}_{nk} \log(\pi_k(w_n, \alpha^{(i+1)})).$$

Q_1 and Q_2 can be maximized separately.

The EM algorithm tends to converge only very slowly and only to a local optimum. To overcome these shortcomings, many variants of the EM algorithm have been proposed in the literature. Two popular versions, which are also implemented in flexmix, are stochastic EM (SEM) and classification EM (CEM). They add an additional step between the expectation and maximization step, in which the estimated a posteriori probabilities are used to assign each observation to exactly one component, either in a stochastic or deterministic way. For further details, refer to the literature.
The number of components for estimation is fixed, but the number of components can be chosen according to the AIC criterion comparing FMMs with different number of components.

D Descriptive Statistics

Table 9: Descriptive statistics of the whole sample

Variable	Minimum	Mean	Maximum
duration of the contract until 01/01/2005 (in years)	0	8.60	35
insurance sum for disability (w/o factor)	0	60,315	230,082
factor	1	3	5
insurance sum for death	0	7,712	51,130
insurance sum for hospital per diem	0	15	65
cost of claims for disability	0	1,930	1,003,953
cost of claims for death	0	59	96,500
cost of claims for hospital per diem	0	4.8	18,630

sample size $n = 957,506$. All figures for sums and costs are in Euro.

References

Abbring, J., Chiappori; P.-A., Pinquet; J., 2003. Moral Hazard and Dynamic Insurance Data. Journal of the European Economic Association 1, 767–820.

Akerlof, G. 1970. The market for lemons: Quality uncertainty and the market mechanism. Quarterly Journal of Economics 84, 488–500.

Autor, D., Duggan, M., 2006. The growth in the social security disability rolls: a fiscal crisis unfolding. NBER Working Paper 12436.

Chiappori, P.-A., Salanié, B., 2000. Testing for asymmetric information in insurance markets. Journal of Political Economy 108, 56–78.

Chiappori, P.-A., Salanié, B., 2012. Asymmetric information in insurance markets: Empirical assessments. In: G. Dionne, editor, Handbook of Insurance, 2nd edition. forthcoming.

Chiappori, P.-A., Jullien, B., Salanié, B., Salanié, F., 2006. Asymmetric information in insurance: General testable implications. Rand Journal of Economics 37, 783–798.

Cohen, A., Siegelman, P., 2010. Testing for adverse selection in insurance markets. Journal of Risk and Insurance 77, 39–84.

Cutler, D., Zeckhauser, R., 2000. The anatomy of health insurance, In: A.J. Culyer and J.P. Newhouse, editors, Handbook of Health Economics.

DeMeza, D., Webb, D., 2001. Advantageous selection in insurance markets. Rand Journal of Economics, 249–262.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM-algorithm. Journal of the Royal Statistical Society B 39, 1–38.

Dionne, G., Doherty, N., Fombaron, N., 2000. Adverse selection in insurance markets. In: G. Dionne, editor, Handbook of Insurance, Chapter 7, pp. 185–244. Kluwer.

Dionne, G., Gourieroux C., and Vanasse, C. (2001). Testing for Evidence of Adverse Selection in the Automobile Insurance Market: A Comment. The Journal of Political Economy 109, 444–453.

Dionne, G., Gourieroux, C., and Vanasse, C., 2006. Informational Content of Household Decisions with Applications to Insurance under Asymmetric Information. in Competitive failures in insurance markets: Theory and policy implications, eds. P. A. Chiappori and C. Gollier, pp. 159–184. Cambridge, MA: MIT Press.

Fang, H., Keane, M., and, Silverman, D., 2008. Sources of Advantageous Selection: Evidence from the Medigap Insurance Market. Journal of Political Economy 116, 303–350.

Finkelstein, A., McGarry, K., 2006. Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market. American Economic Review 96, 938–958.

Finkelstein, A., Poterba, J., 2006. Testing for adverse selection with 'unused observables'. NBER Working Paper 12112.

Fruehwirth-Schnatter, S., 2006. Finite Mixture and Markov Switching Models. Springer-Verlag.

Gan, L., Huang, F., Mayer, A., 2011. A simple test of private information in the insurance markets with heterogeneous insurance demand. NBER Working Paper 16738.

Gesamtverband der Deutschen Versicherungswirtschaft (GDV), 2010. Statistisches Taschenbuch der Versicherungswirtschaft 2010.

Gourieroux, C., Monfort, A., Renault, E., Trognon, A., 1987. Generalised residuals. Journal of Econometrics 34, 5–32.

Greene, W., 2008. Econometric Analysis. 7th Edition.

Gruber, J., Kubik, J., 1997. Disability insurance rejection rates and the labor supply of older workers. Journal of Public Economics 64, 1–23.

Gruen, B., Leisch, F., 2007. Fitting finite mixtures of generalized linear regressions in R. Computational Statistics & Data Analysis 51, 5247–5252.

Gruen, B., Leisch, F., 2008. FlexMix Version 2: Finite mixtures with concomitant variables and varying and constant parameters. Journal of Statistical Software 28, 1–35, 2008. URL http://www.jstatsoft.org/v28/i04/

Hendren, N., 2013. Private Information and Insurance Rejections. Econometrica, forthcoming.

Henry, M., Kitamura, Y., Salanie, B., 2013. Partial Identification of Finite Mixtures in Econometric Models. Quantitative Economics, forthcoming.

Holmström, B., 1979. Moral hazard and observability. The Bell Journal of Economics 10, 74–92.

Kalbfleisch, J.D., Prentice, R.L., 2002. The Statistical Analysis of Failure Time Data. Wiley.

Klein, J.P., Moeschberger, M.L., 2003. Survival Anaylsis. 2nd edition. Springer-Verlag.

Leisch, F., 2004. FlexMix: A general framework for finite mixture models and latent class regression in R. Journal of Statistical Software 11, 1–18.
URL http://www.jstatsoft.org/v11/i08/

McLachlan, G.J., Peel, D., 2000. Finite Mixture Models. Wiley.

Puelz, R., Snow, A., 1994. Evidence on adverse selection: Equilibrium signaling and cross-subsidization in the insurance market. Journal of Political Economy 102, 236–257.

Rothschild, M., Stiglitz, J. E., 1976. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. The Quarterly Journal of Economics 90, 630–649.

Sass, A.K., 2008. Unfallgeschehen in Deutschland: Auswertung des telefonischen Gesundheitssurveys 2004 (Accidents in Germany) Deutsches Ärzteblatt 105(36): 604–8.

Shavell, S., 1979. On moral hazard and insurance. The Quarterly Journal of Economics 93, 541–562.

Spindler, M., Winter, J., Hagmayer, S., 2011. Asymmetric information in automobile insurance: Evidence from Germany. Journal of Risk and Insurance, forthcoming.

Statistisches Bundesamt (edior), 2011. Mikrozensus - Fragen zur Gesundheit - Kranke und Unfallverletzte.

Su, L., Spindler, M., 2011. Nonparametric testing for asymmetric information. Journal of Business and Economic Statistics 31, 208–225.

Winter, R., 2000. Optimal insurance under moral hazard. In: G. Dionne, editor, Handbook of Insurance, Chapter 6, pp. 155–184. Kluwer.

Yee, T.W., 2009. VGAM: Vector Generalized Linear and Additive Models. R package version 0.7-9, http://www.stat.auckland.ac.nz/ yee/VGAM.