Essays in Econometrics

Inaugural–Dissertation
zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig–Maximilians–Universität München

2011

vorgelegt von
Martin Spindler

Referent: Prof. Dr. Joachim Winter
Korreferent: Prof. Dr. Gebhard Flaig
Promotionsabschlussberatung: 16. Mai 2012

Acknowledgement

First, I would like to thank my advisor Joachim Winter for his enormous efforts and constant support for both research and organizational issues. He was always willing to take time and to listen to my problems.

I am also grateful to Liangjun Su. The cooperation with him was a great experience. Working jointly with him was a fruitful source of motivation and learning.

Furthermore I am indebted to Bernard Salanié who accompanied my research from the early beginnings. He helped me out with providing a data set from the French car insurance to get started. Moreover, he enabled me a research stay at the Columbia University which had an immense influence on both my academic and personal development.

I am also grateful to Andreas Richter who agreed to serve on the thesis committee and showed much interest in my research, amongst others by inviting me to his seminars and giving valuable comments and support.

I thank Gebhard Flaig for agreeing to act as second reviewer.

At the beginning of my research project there was the idea to test the theory of asymmetric information with German insurance data. After one year of writing reams of research proposals without getting data and when I was actually close to desperation, two lucky events occurred: The insurance companies HUK Coburg and AXA showed interest in my research and agreed on a cooperation. Therefore I am very grateful for their dedication. For the HUK Coburg it is by name Jörg Rheinländer and Steffen Hagmayer who also became a coauthor and prepared the data set excellently and answered all my questions patiently. Jochen Kneiphof and Li Zhang from AXA promoted my research on the accident insurance by providing data and giving valuable input.

Some calculations were done by using R. Bettina Gruen gave me an introduction to the package flexmix and its operating mode with extraordinary patience. Thomas Yee added resp. modified functions to the package VGAM upon my request. I appreciate their help highly.

As a member of the Munich Graduate School of Economics I would also like to mention

gratefully its former director Sven Rady for enabling me the participation in conferences.
Last, but not least I thank Felix, Maximilian and Thomas for valuable discussions about statistical, mathematical and computational problems.
Finally, I gratefully acknowledge financial support from the Deutsche Forschungsgemeinschaft through GRK 801, from the German Academic Exchange Service (DAAD) and in particular the GENEVA Association.

Contents

List of Tables

Chapter 1

Introduction

The concept of asymmetric information which comprises two distinct phenomena, notably moral hazard and adverse selection, establishes the core of microeconomics. The whole subfield of contract theory is built on the notion of asymmetric information. The importance is also stressed by the award of Nobel Prize in Economics to pioneers in this field in the year 2001.

Adverse selection describes the situation in which parties have different information before a contract is signed. This idea was introduced in the highly celebrated article Akerlof (1970). Akerlof showed that informational asymmetries might influence the functioning of markets or may even lead to market failure. His leading example was the second-hand car market where potential sellers have an advantage in assessing the quality of their car.
Moral hazard characterizes problems which arise after a contract is signed and results from different levels of information between parties, e.g., because one party can observe the outcome of certain events while the other is not able to do this. This idea was worked out in Holmström (1979).

Since its introduction, asymmetric information has successfully been applied (mostly theoretically) amongst others to public finance, finance, labor economics, personnel economics and insurance economics and the theory has been extended in many directions.
The term "adverse selection" itself originated in the context of insurance. Insurance markets have become one of the earliest applications. Arrow (1963), Pauly (1974) and Rothschild and Stiglitz (1976) analyzed insurance markets under the assumption that insurees have more information about their risk than the insurance companies have. In this case in equilibrium a menu of contracts is offered and "good" risks choose a partial insurance while "bad" risks choose full coverage at a higher price per unit of insurance. After an insurance contract is signed the incentives for careful behavior of an insured person are reduced as the individual

conduct cannot be observed by the insurance company in many cases and possible damages are borne by the insurance company. Therefore moral hazard is also important in insurance markets as first pointed out by Shavell (1979).

In economics in general and in insurance economics in particular the theory of asymmetric information has been developed at a rapid pace. Surveys for the field of insurance are Dionne et. al. (2000) and Winter (2000). But in contrast to the theoretical development the empirical and econometric analysis lagged behind and to this point we would like to contribute.
One problem in the empirical analysis of contract theory is that either suitable data sets for the question at hand do not exist or - if they exist - are often not available for research. But insurance markets have proved a fruitful and productive field for empirical studies. In most cases insurance contracts are standardized. Insurance companies store the relevant information which is necessary to test the theoretical predictions and they have large data bases in order to run statistical analysis. Cohen and Siegelman (2010) give a recent overview of testing for asymmetric information in insurance markets. One central question is, if insurees really have a better assessment of their risk than insurance companies. This question and the question if asymmetric information is prevalent in a market are important for contractual, institutional and regulatory issues. Problems for which knowledge about the extent of asymmetric information is decisive are, e.g., how to design social insurance, either to introduce a compulsory public insurance in the case of market failure or to relinquish it to the private sector. An other important question concerns regulation. In order to assess the effects of prohibition to use certain variables for risk classification on the market outcome, an analysis of the extent of adverse selection is indispensable. A recent example is the ban to base the insurance rate on sex. Such interventions might even lead to a break down of markets if adverse selection is dominant. Recent results indicate that there is no general answer to this question and it seems that it depends on the concrete market resp. risk under consideration and the detailed contractual and institutional design.

Testing for asymmetric information does not have a long tradition in economics. The basic models of asymmetric information predict a positive correlation between risk and coverage. Chiappori and Salanié (2000) introduce tests for this positive correlation. This approach was widely applied in other studies. One drawback of parametric approaches is that they are vulnerable with respect to the functional and distributional specification. Chiappori and Salanié (2000) also apply a nonparametric test for which all variables have to be binary. Such a transformation might lead to a loss of information. Therefore a general nonparametric test for asymmetric information is beneficial.
One problem connected with tests for positive correlation is that also the result of zero cor-

relation might be compatible with the existence of selection on risk, i.e., adverse selection. DeMeza and Webb (2001) show that if insurees differ in risk aversion (preferences) and if additionally risk averse individuals are ,e.g., more cautious then selection on preferences might superimpose the selection on the risk type and might lead to a negative or zero correlation despite adverse selection. Therefore the interpretation of the positive correlation test is not unambiguous. There are several ways to circumvent this problem. Chiappori et al. (2006) introduce a generalized positive correlation property which even holds if the agents have different preferences. Another way to control for differences in preferences is to use unused observables as proposed by Finkelstein and Poterba (2006). Both tests have high demands concerning the data. They require either information about the premium or the existence of variables which are contained in the data set but not used for pricing resp. risk classification (so called unused observables). An other possible procedure with lower data requirements is to use finite mixture models to account for unobserved heterogeneity. In order to test for asymmetric information with them no more information is required than with the standard test procedures.

In chapter 2 we tackle two distinct problems: one methodological and one empirical.
The basic models of asymmetric information predict a positive correlation between risk and coverage and there are several ways to test for this property. Procedures have been proposed by Chiappori and Salanié (2000), Dionne et al. (2001) and Kim et. al. (2009). One natural question is if all these tests deliver consistent results, not in a statistical sense but in the sense of answering the underlying question in the same way. These tests build on different principles resp. different translations of the formal definition of asymmetric information into a statistical framework. We show that all tests deliver robust results and therefore the choice of a particular method is only of minor importance.
Moreover, the literature has reached a consensus that asymmetric information is not prevalent in the automobile insurance.[1] Our analysis shows that this conclusion is not correct in general and that the arrangement and organization of the markets and the contracts have a great influence on the fact if insurees have an informational advantage and possibly can use this advantage. Because of a special arrangement of the car insurance in Germany we can show that the extent of asymmetric information depends on the kind of risk which is precisely covered. Therefore we stress the importance of institutional and contractual conditions in detail which finally enable or disable the insureds to use possible asymmetries in the information structure.

In chapter 3 a nonparametric test for asymmetric information is proposed and applied to French automobile data. The test is based on an alternative interpretation what asym-

[1] For a detailed literature review we refer to chapter 2.

metric information resp. its absence means. The absence of asymmetric information means that the choice of a contract Y (discrete variable) provides no information for predicting the "performance" variable Z (discrete or continuous, e.g., the number of claims or the sum of reimbursements), conditional on the vector X of all exogenous variables (discrete and continuous). Therefore we can transform the problem of testing the absence of asymmetric information into a test for conditional independence: $F(Z|X,Y) = F(Z|X)$ almost surely (a.s.) where, e.g., $F(Z|X,Y)$ denotes the conditional cumulative distribution function (CDF) of Z given (X,Y). We propose a nonparametric test statistic to test the conditional independence of Z and Y given X. We show that the test statistic is asymptotic normally distributed under the null hypothesis of conditional independence (or absence of asymmetric information) and diverges to infinity in the presence of conditional dependence (or asymmetric information).
Parametric tests are fragile to both *functional* and *distributional* form misspecification which are a severe problem in this field. For example, in the automobile insurance market it is common knowledge that the age of the driver has a nonlinear effect on the probability of an accident, but such a nonlinear effect has rarely been taken into account in the literature. For another example, the error term in the binary model for modeling the choice of an insurance contract may not be either normally or logistically distributed, and tests for asymmetric information based on the probit or logit model can therefore yield misleading conclusions in the case of incorrect distributional specification. Our test can control for both for arbitrary *functional* and *distributional* forms.
We then apply our test to a French automobile insurance data set and compare our testing results with the results found in the literature.

In the last part of the dissertation, which consists of two chapters, we pursue two goals: one is to deliver results for the accident insurance which covers elementary risks of individuals. The other is to control for unobserved heterogeneity which might lead to a zero correlation of risk and coverage although there is selection on risk. DeMeza and Webb (2001) show, as mentioned above, that also a zero or negative correlation is compatible with adverse selection and this complicates the interpretation of results.
In order to deal with this problem there are two possible solutions: the "unused"observables test introduced by Finkelstein and Poterba (2006) or finite mixture models to account explicitly for unobserved heterogeneity. As the data set we analyze contains variables which are not used for risk classification - a rare situation - we can apply the test with unused observables. The basic idea is that if there is a variable which is not used for risk classification but simultaneously influences the choice of coverage and is a predictor for risk then there is asymmetric information.

Moreover, we will also apply mixtures of bivariate probit models to account for heterogeneity. A first application of finite mixture models to test for asymmetric information is found in Gan et. al. (2011).

First, we analyze the data set with the standard parametric and a nonparametric test. In order to be able to compare the results, we apply all tests to the same subsample. Because of computational limitations the nonparametric test can only be applied to relatively small subsamples. We find evidence for the existence of asymmetric information for small and middle insured sums. For high values the evidence is not so clear.

In the second part, we try to control for the unobserved heterogeneity by applying the methods mentioned above. We find asymmetric information even when controlling for unobserved heterogeneity while the extent varies with the chosen insured sum.

As the dissertation was written in a cumulative way and has its origin in several papers, some sections of the chapters are similar. But to avoid confusion and unnecessary cross references they have been retained.

References

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

Arrow, K.J., 1963. Uncertainty and the welfare economics of medical care. American Economic Review 53, 941-973.

Chiappori, P.-A., Salanié, B. 1997. Empirical contract theory: the case of insurance data. European Economic Review 41, 943-950.

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

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.

de Meza, D., Webb, D., 2001. Advantageous selection in insurance markets. Rand Journal of Economics, 249-262.

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

Dionne, G., Gourieroux C., Vanasse, C. 2001. Testing for evidence of adverse selection in the automobile insurance market: A comment. The Journal of Political Economy 109, 444-453.

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

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.

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

Kim, H., Kim, D., Im, S., Hardin, J. W., 2009. Evidence of asymmetric information in the automobile insurance market: dichotomous versus multinomial measurement of insurance coverage. The Journal of Risk and Insurance 76, 343-366.

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.

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

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

Chapter 2

Asymmetric Information in the Market for Automobile Insurance: Evidence from Germany*

2.1 Introduction

Since Akerlof (1970), the consequences of asymmetric information, in particular adverse selection and moral hazard, have been explored in a vast body of research. The initial gap between the theoretical developments and empirical studies of asymmetric information has recently become narrower. In particular, insurance markets have proved a fruitful and productive field for empirical studies, for two reasons: First, insurance contracts are usually highly standardized; they can be described completely by a relatively small set of variables and data on the insured person's claim history, i.e., the occurrence of claims and the associated costs, is stored in the database of an insurance company. Second, insurance companies have hundreds of thousands or even millions of clients and therefore the samples are sufficiently large to conduct powerful statistical tests. The markets for automobile insurance, annuities and life insurance, crops insurance, as well as long-term care and health insurance provide large samples of standardized contracts for which performances are recorded and are well suited for testing the theoretical predictions of insurance theory. Chiappori and Salanié (1997) provide a detailed justification for using insurance data to test contract theory. Cohen and Siegelman (2010) present a comprehensive overview of approaches for testing for adverse selection in insurance markets, covering a large number of empirical studies in different insurance branches.

*This chapter is based on joint work with Joachim Winter and Steffen Hagmayer.

In statistical terms, the notion of asymmetric information implies a positive (conditional) correlation between coverage and risk. Several different methods how to test for asymmetric information have been proposed in the literature. We present the most important and powerful (parametric) tests and apply them to a German car insurance data set. Our study contributes to the existing literature in several respects. We present the first study analyzing the German car insurance market. The German car insurance market is the largest in Europe and therefore for many insurance companies the most important sales market for their insurance policies. We had unique access to the data set of one of the largest insurance companies in the field of automobile insurance in Germany.

Second, the literature has reached consensus that asymmetric information is not prevalent in the automobile insurance. Our analysis shows that his conclusion is not correct in general and that the arrangement and organization of the markets and contracts have a great influence on the fact if insurees have an informational advantage and possibly can use this. Because of a special arrangement of the car insurance in Germany we can show that the extent of asymmetric information depends on the kind of risks which are precisely covered. Therefore we stress the importance of institutional and contractual conditions in detail which finally enable or disable the insureds to use possible asymmetries in the information structure.

Third, several tests for asymmetric information have been proposed in the literature. Chiappori and Salanié propose tests for the positive correlation property, Dionne et. al. (2001) use a two stage approach and Kim et. al. (2009) modify a multinomial approach. Most studies only apply a selection of these tests. We apply all tests on the same data set and see that they give consistent results, not in a statistical sense but in the sense of answering the underlying question in the same way. These tests build on different principles resp. different translations of the formal definition of asymmetric information into a statistical framework. We show that all tests deliver robust results and therefore the choice of a particular method is only of minor importance.

Finally, by applying the framework of Chiappori et al. (2006) we can also test whether consumers know their loss distribution, if in this market the non-increasing profit assumption holds, i.e., that contracts with higher coverage earn not higher profits, and, most importantly, we can test if some form of generalized positive correlation property holds, which also allows for differences in risk preferences. The first two statements are interesting on their own although they serve as assumptions for the last one. One important practical question in insurance economics is if insurees can correctly estimate their loss distribution or if the overestimate or underestimate their risk. We test this in section 6. Another important issue in insurance markets is the question about the market structure. In section 6 we can show that the nonincreasing profit condition holds in this market and in this period, which indicates that there is strong competition among insurance firms, which is in line with descriptions of this particular market that we discuss in section 3 suggest.

The rest of the paper is structured as follows. Section 2 outlines the theory of asymmetric information and summarizes the empirical literature relevant for our paper. Section 3 describes the arrangement of car insurance in Germany and the structure of the German automobile insurance market. In section 4, we describe the data set. In section 5, we review briefly the parametric tests used in this field and present the results for our data set. In section 6, we introduce the generalized positive correlation property and some related tests and present the results. Final remarks and conclusions are contained in Section 7.

2.2 Asymmetric Information in Insurance Markets: Theory and Evidence

In their seminal paper Rothschild and Stiglitz (1976) introduce the notion of adverse selection in insurance markets that has since then been extended in many directions.[1] In the basic model, the insureds have private information about the expected claim, exactly speaking about the probability that a claim with fixed level occurs, while the insurers do not have this information. Thus there are two groups with different claim probabilities, the "bad" and "good" risks. The agents have identical preferences which are moreover perfectly known to the insurer. Additionally, perfect competition and exclusive contracts are assumed. Exclusive contracts mean that an insuree can buy coverage only from one insurance company. This allows firms to implement nonlinear (especially convex) pricing schemes which are typical under asymmetric information. Under this setting insurance companies offer a menu of contracts in equilibrium: a full insurance which is chosen by the "bad" risks and a partial coverage which is bought by the "good" risks. In general, contracts with more comprehensive coverage are sold at a higher (unitary) premium.

Clearly, one expects a positive correlation between "risk" and "coverage" (conditional on observables). Since the assumptions in the Rothschild and Stiglitz model are very simplistic and normally not fulfilled in real applications, an important question to address is how robust this coverage-risk correlation is. Chiappori et al. (2006) show that the positive correlation property extends to much more general models under a suitable defined notion. Especially the notion of positive correlation is generalized in this context. Under competitive markets

[1]For a detailed survey on adverse selection and the related moral hazard problem, see Dionne, Doherty and Fombaron (2000) and Winter (2000), respectively.

this property is also valid in a very general framework entailing heterogeneous preferences, multiple level of losses, multidimensional adverse selection plus possible moral hazard and even non-expected utility theory. In the case of imperfect competition some form of positive correlation must hold if the agent's risk aversion becomes public information. In the case of private information the property does not necessarily hold (c.f. Jullien et al. (2007)).

While adverse selection concerns "hidden information", moral hazard deals with "hidden action". Moral hazard occurs when the expected loss (accident probability or level of damage) is not exogenous, as assumed in the adverse selection case, but depends on some decision or action made by the subscriber (e.g., effort or prevention) which is neither observable nor contractible. A higher coverage leads to decreased efforts and therefore to a higher expected loss. Therefore moral hazard also predicts a positive correlation between "coverage" and "risk".

Although both phenomena lead to a positive risk-coverage correlation, there is one important difference: under adverse selection the risk of the potential insuree affects the choice of the contract, whereas under moral hazard the chosen contract influences the behavior and therefore the expected loss. So there exists reversed causality in both cases.[2]

In sum, the theory of asymmetric information[3] predicts a positive correlation between (appropriately defined) "risk" and "coverage" which should be quite robust.

To proceed, it is worth mentioning that to test for asymmetric information, the researcher needs to have access to the same information which is also available to the insurer and used for pricing. The theory of adverse selection predicts that the insurance company offers a menu of contracts to indistinguishable individuals. Individuals are (ex ante) indistinguishable for the insurer if they share the same characteristics. Therefore the positive risk-coverage correlation is valid only conditional on the observed characteristics. Different groups of observable equivalent individuals are offered different menus of contracts with different prices according to their risk exposure.[4] Only within each class are the mechanisms described above valid.

Despite the scarcity of data sets in empirical contract theory, the automobile insurance market has been analyzed extensively. Amongst others, automobile insurance markets in France (Salanié and Chiappori (2000, 2006) and Richaudeau (1999)), Israel (Cohen (2005)), Canada (Dionne et al. (2001)), Korea (Kim et al. (2009)), Japan (Saito (2009)) and the Netherlands (Zavadil (2011)) have been analyzed. In one of the first studies, Puelz and Snow

[2]To disentangle moral hazard from adverse selection is an important problem in the empirical literature. The first attempt is Dionne et al. (2004). An overview over different possible strategies for dealing with this problem can be found in Cohen and Siegelman (2010).

[3]It seems that in the empirical insurance literature adverse selection is more stressed than the moral hazard aspect which only receives minor attention, see, e.g., Cohen and Siegelman (2010).

[4]For the theory of risk classification under asymmetric information see Crocker and Snow (2000).

(1994) confirm the existence of asymmetric information, but Dionne et al. (2001) show that this might be due to misspecification of their model. In general, there is a tendency to confirm absence of asymmetric information, e.g., Salanie and Chiappori (2000), Kim et al. (2009), Dionne et al. (2001) and Zavadil (2011). Only in the market of Israel asymmetric information is found for experienced drivers but this can be contributed to a special feature of this market. Insurance companies in Israel cannot gather information about the driving history of their new customers. This gives consumers who change their insurer some advantage. The evidence for asymmetric information found by Kim et al. (2009) seems to be relatively weak.

2.3 Automobile Insurance in Germany

Like in many other countries, a third-party vehicle insurance is mandatory for all vehicles in Germany. This is the so called "KFZ-Haftpflicht". This is a liability insurance that covers damage inflicted to other drivers and their cars. Moreover, insurance companies offer additional non-compulsory coverage where two different types must be distinguished which cover own damages. The first one is the so called "Teilkasko"(TK). This type covers own damages and losses caused by theft, natural disasters (storm, hail, lightning strike, flood), collusion with furred game and so on. The second type is the "Vollkasko"(VK). It covers accidental damage on the own car, even if caused by oneself, and damages caused by vandalism of strangers. For both types the insuree can choose a deductible.

In the German car insurance there is also a so called "Schadendsfreiheitsrabatt"(SFR), a uniform experience rating system which applies only to the "Vollkasko", but not to damages due to the "Teilkasko". The number of years without accident are counted separately for the "Haftpflicht "and the "Vollkasko"and according to these numbers every insuree is divided into a certain class ("Schadensfreiheitsklasse"(SFK)). With every class a certain coefficient b_t is associated which is a proxy for past experience. At any date/year t, the premium is defined as the product of a basis amount and this coefficient. The basic amount can be defined freely by the insurance companies according to their risk classification but cannot be related to past experience. Suppose, the bonus coefficient is b_t at the beginning of the tth period. Then the occurrence of an accident during the period leads to a categorization into another SFK class and ,e.g., an increase of 25 percent at the end of the period (i.e., $b_{t+1} = 1.25b_t$), whereas an accident-free year leads to a reduction of the coefficient according to the new class. Additionally, the coefficient is to be restricted to lie between 230 % and 30 % in the "Haftpflicht"and between 125 % and 30 % in the "Vollkasko".

The basis amount of the premium is calculated according to different risk classes. Due

Table 2.1: An overview of the German car insurance market in 2008

	"Haftpflicht"	"Teilkasko"	"Vollkasko"
number of insured cars in million	39.69	12.6	20.76
number of claims in million	2.57	1.30	3.43
claims expenditure in billion Euro	9.22	0.95	5.00
average claim in Euro	3,600	730	1,460

source: GDV (2010).

to variables like age, sex, profession, area, etc., the insurees are divided into different risk classes which should reflect their accident probabilities, and the premium to be paid is then determined.

In 2009, the size of German motor insurance markets was about 20 billion Euro. 39.69 million cars were covered by "Haftpflichtversicherung" which is compulsory for every car owner; this is thus the total number of registered cars. A detailed survey of key figures for the year 2008 is given in Table 2.1.

In 2009 the premium income was 20,057 million Euro and expenditures for claims were 19,420 million Euro. There are 104 companies in the market competing for contracts, a wider offer than in many other countries in western Europe. A very important statistic for insurance companies is the so called "combined ratio", cost and claims divided by the premium income. This figure was in the last years about 100 %, in 2008 and 2009 the industry average was slightly above 100 %. A combined ration over 100 indicates that the insurer is making an underwriting loss. The reason is that in the last five years there was a price battle in this market with cutting rates in each year leading to insurance rates similar to the level of those of the early 1980s (see Bloomberg (09/06/2010)). Between 2004 and 2009, the average premium decreased by 15.9 % (cf. GenRe (2010)). A detailed analysis of the German car insurance and the prevailing price war in the last years is given in GenRe (2010). While most insurance markets show a tendency towards oligopolies, in the last years the German car insurance market was very competitive, close to perfect competition as the figures above and recent market surveys indicate.[5] In section 6 we also test for the non increasing profit hypothesis which holds in most settings of competitive markets but is more general than a simple null-profit condition.

[5] GenRe (2010), Bloomberg (11/13/2009), Reuters (01/28/2009), Reuters (05/17/2010) and Handelsblatt (09/01/2010) give recent information for the developments on the German car insurance market, especially for the existing price war and the hard competition.

2.4 The Data Set

For our analysis we had access to the database of the insurance contracts of a major company in Germany. We conducted separated analysis for the "Teilkasko"(TK) and "Vollkasko"(VK) due to the different scopes of indemnity and liability rules. Besides, we analyzed both the whole portfolio of contracts and a subsample restricted to young drivers[6]. The concentration on beginners enables us to rule out learning effects which might arise in time on the side of the insuree or the insurance company. As the database is too voluminous we restrict our analysis in both cases to random samples.[7] We use data comprehending the year 2009. The data set contains information about each contract in the sample for a full contract year. The sample size in the TK is $n = 5,321$ for the whole portfolio and $n = 5,647$ for the beginners, in the VK $n = 7,200$ resp. $n = 6,466$ for the beginners. The level of deductibles in the TK are 0, 150, 300 and 500 Euro and 0, 300, $1,000$ and $2,500$ Euro in the VK. As $2,500$ Euro is very seldom chosen we omit this level of deductible. In the TK case we use the number of claims exceeding 500 Euro and in the VK the number of claims exceeding twice the highest level of deductible, i.e., $2 \cdot 1,000 = 2,000$ Euro as a measure for the ex post risk. In the TK there is no incentive not to claim accidents as there is no bonus / malus coefficient. In the VK case there is an incentive not to claim all accidents as accidents lead to a worsening of the the risk classification and therefore to a higher premium. But we argue that claims higher than $2,000$ Euro are filed in any case. In general, as accidents below the deductible are not reported to the insurance company and cannot be observed we have to restrict to claims being higher than certain thresholds as mentioned above. Otherwise also in the absence of asymmetric information a positive correlation would be possibly detected.

A detailed statistical analysis of the chosen deductibles and the number of accidents according to the chosen deductibles for both the whole portfolio and the novice drivers broken down into Teilkasko and Vollkasko is provided in Tables 2.2–2.5.

As our data set stems from the company with one of the highest market shares and is diversified across regions in Germany, across all occupations and ages it can be assumed that the structure of our data is representative for the the whole population and therefore our results can be generalized to the whole market.

---

[6]In actuarial science the expression "young drivers"refers not to the age of the driver but to the driving experience.

[7]For the analysis of the whole portfolio the random sample was drawn from the set of all contracts which were signed after January 1, 2007. As there are changes in the product menu and contract conditions from time to time, thereby we can secure that the contracts are really comparable and differ only in the chosen deductible.

Table 2.2: Number of accidents according to the choice of deductible in the TK for the whole portfolio

level of deductible / number of accidents	0	1	2	sum
0	552 (0.914)	50 (0.083)	4 (0.003)	604
150	4349 (0.953)	208 (0.046)	7 (0.001)	4564
300	107 (0.964)	4 (0.036)	0 (0.000)	111
500	40 (1.00)	0 (0.000)	0 (0.000)	40

Notes: Figures in brackets denote the relative frequency of the number of accidents for the corresponding level of deductible.

Table 2.3: Number of accidents according to the choice of deductible in the VK for the whole portfolio

level of deductible / number of accidents resp. share	0	1	2	sum
0	54 (0.982)	1 (0.018)	0 (0.000)	55
150	486 (0.972)	14 (0.028)	0 (0.000)	500
300	5182 (0.977)	119 (0.022)	2 (0.001)	5303
500	962 (0.969)	31 (0.031)	0 (0.000)	993
1000	129 (0.985)	2 (0.015)	0 (0.000)	131

Notes: Figures in brackets denote the relative frequency of the number of accidents for the corresponding level of deductible.

Table 2.4: Number of accidents according to the choice of deductible in the TK for the novice drivers

level of deductible / number of accidents	0	1	2	sum
0	377 (0.887)	44 (0.103)	4 (0.010)	425
150	4723 (0.940)	288 (0.057)	15 (0.003)	5026
300	148 (0.961)	6 (0.039)	0 (0.000)	154
500	40 (1.00)	2 (0.000)	0 (0.000)	40

Notes: Figures in brackets denote the relative frequency of the number of accidents for the corresponding level of deductible.

Table 2.5: Number of accidents according to the choice of deductible in the VK for the novice drivers

level of deductible / number of accidents resp. share	0	1	2	3	sum
0	3 (1.000)	0 (0.000)	0 (0.000)	0 (0.000)	13
150	205 (0.995)	1 (0.005)	0 (0.000)	0 (0.000)	206
300	4619 (0.964)	167 (0.035)	6 (0.001)	1 (0.000)	4793
500	1194 (0.954)	56 (0.045)	2 (0.001)	0 (0.000)	1252
1000	129 (0.960)	8 (0.040)	0 (0.000)	0 (0.000)	202

Notes: Figures in brackets denote the relative frequency of the number of accidents for the corresponding level of deductible.

2.5 Testing for Adverse Selection and Moral Hazard

2.5.1 Statistical Procedures

In this section, we present several different methods for testing asymmetric information. In an econometric sense we want to test if there is a positive correlation between risk and coverage.[8] X denotes the exogenous variables which are used for risk classification by the insurance company, Y the chosen contract, e.g., the chosen deductible, and Z measures the risk. The risk is measured as "ex post risk", e.g., by the number of accidents or the caused damage payments by the insuree. An index i refers to a certain individual resp. contract which is omitted if there is no confusion.

2.5.1.1 Unrelated probit regressions

The first approach is to define two probit models, one for the choice of the coverage Y_i (either compulsory/basic coverage or comprehensive coverage) and the other for the occurrence of an accident Z_i (either no accident being blamed for or at least one accident with fault):

$$\begin{cases} Y_i = \mathbf{1}(X_i\beta + \varepsilon_i > 0) \\ Z_i = \mathbf{1}(X_i\gamma + \eta_i > 0) \end{cases} \tag{5.1}$$

where ε_i and η_i are independent standard normal errors, and β and γ are coefficient vectors (as columns). The row vector X_i denotes the covariates of individual i. First these two

[8] We will concentrate on parametric tests which are well established in this field and can be implemented by most statistical software packages. Spindler and Su (2011) present a nonparametric test for asymmetric information.

Table 2.6: Description of the variables

name	description	number of categories "Teilkasko"	number of categories "Vollkasko"
commitment to work-shop	yes/no	2	2
profession	categorial	9	9
region	categorial	6	4
type of vehicle	categorial	6	6
no-claims bonus	bonus / malus coefficient; categorial	–	8
kilometers per year	categorial	9	9
age of car when being bought	categorial	8	7
lodging of the car over night combined with housing	categorial	12	12
driver	age of the driver and potential drivers – categorial	18	25
keeper of the car	categorial	5	5
payment method	categorial	6	6
bonus	yes/no	2	2
protection against upgrading	yes /no, in the case of an accident the no-claims bonus is not raised	–	2
deductible	different possible values	4	5
number of accidents	discrete		
payment for damages	continuous		

Note: For the young drivers the number of categories is slightly different, i.e., smaller, as they have a little bit lower variation in their characteristics.

probit models are estimated independently and then the generalized residuals $\hat{\varepsilon}_i$ and $\hat{\eta}_i$[9] are calculated. These are required for the following test statistic

$$W_n = \frac{\left(\sum_{i=1}^{n} \hat{\varepsilon}_i \hat{\eta}_i\right)^2}{\sum_{i=1}^{n} \hat{\varepsilon}_i^2 \hat{\eta}_i^2}. \tag{5.2}$$

Under the null of conditional independence, $\text{cov}(\varepsilon_i, \eta_i) = 0$ 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.

2.5.1.2 Bivariate Probit regression

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 hast to be estimated, and then to test whether $\rho = 0$ or not. In order to test this hypothesis the Wald-, Score- oder LR-test can be used.

2.5.1.3 Two-stage regressions

Multinomial approach (Kim et al. (2009)) Depending on the number of categories of the choice variable Y this procedure varies a little bit. In the case of a dichotomous Y in the first stage a bivariate probit of the choice of contract on the exogenous variables, i.e., the variables used for risk classification, is run. The probit equation is of the form

$$Y_i = \mathbf{1}(X_i\beta + \varepsilon_i > 0) \tag{5.3}$$

with ε_i iid and standard normal. Then the generalized residuals $\hat{\varepsilon}_i$ are obtained:

$$\hat{\varepsilon}_i = \frac{\phi(X_i\hat{\beta})}{\Phi(X_i\hat{\beta})(1-\Phi(X_i\hat{\beta}))}\left[Y_i - \Phi(X_i\hat{\beta})\right], \tag{5.4}$$

where ϕ and Φ are the density and cumulative distribution functions of the standard normal distribution and $\hat{\beta}$ is the estimated coefficient vector. As an unexplained probability of making a corresponding coverage choice, $\hat{\varepsilon}_i$ captures the extent of private information in the binary choice of Y_i, conditional on the observables.
In the case of more than two categories of Y_i an ordered multinomial choice model is applied.

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

For example Y_i is equal to 0 if policyholder i chooses no optional coverage (i.e., liability only), 1 for some optional coverage and 2 for all optional coverage / full coverage. For the kind of insurance we will analyze (separately for Teil- and Vollkasko) Y_i is equal to 0 if someone chooses the highest possible deductible, equal to 1 if someone chooses some medium deductible and 2 for no deductible.
A multinomial choice model is given by

$$Y_i = \begin{cases} 0 & \text{if} \quad Y_i^* \leq \mu_1 \\ 1 & \text{if} \quad \mu_1 < Y_i^* \leq \mu_2 \\ 2 & \text{if} \quad Y_i^* \leq \mu_2 \end{cases} \tag{5.5}$$

where Y_i^* denotes a latent variable representing the policyholders utility associated with insurance coverage, and μ_1 and μ_2 are unknown thresholds for observed categories. The ordered multinomial choice model can be estimated using an ordered probit regression.

With an ordered multinomial variable $Y_i \in \{0, 1, 2\}$ the above procedure is not directly applicable. In order to obtain the unexplained probabilities equivalent to the one in the binary choice model the choice of three contracts is split up in two binary choices.[10] Therefore they define two auxiliary variables Y_i^1 and Y_i^2. $Y_i^1 = 0$ if $Y_i = 0$ and $Y_i^1 = 1$ if $Y_i \in \{1, 2\}$. Accordingly, $Y_i^2 = 0$ if $Y_i \in \{0, 1\}$ and $Y_2^1 = 1$ if $Y_i = 2$. Then the generalized residuals are calculated by

$$\hat{\varepsilon}_i^1 = \frac{\phi(X_i\hat{\beta} - \hat{\mu}_1)}{\Phi(X_i\hat{\beta} - \hat{\mu}_1)(1 - \Phi(X_i\hat{\beta} - \hat{\mu}_1))} \left[Y_i^1 - \Phi(X_i\hat{\beta} - \hat{\mu}_1) \right] \tag{5.6}$$

$$\hat{\varepsilon}_i^2 = \frac{\phi(X_i\hat{\beta} - \hat{\mu}_2)}{\Phi(X_i\hat{\beta} - \hat{\mu}_2)(1 - \Phi(X_i\hat{\beta} - \hat{\mu}_2))} \left[Y_i^2 - \Phi(X_i\hat{\beta} - \hat{\mu}_2) \right] \tag{5.7}$$

with $\hat{\beta}$ the estimated coefficient vector and $\hat{\mu}_1$, $\hat{\mu}_2$ the estimated thresholds for observed categories.
$\hat{\varepsilon}_i^1$ and $\hat{\varepsilon}_i^2$ can be interpreted as the unexplained probabilities associated with Y_i^1 and Y_i^2. For example, $\hat{\varepsilon}_i^1$ estimates the unexplained probability of choosing any optional coverage ($Y_i \in 1, 2$) over no optional coverage ($Y_i = 0$). $\hat{\varepsilon}_i^2$ can be interpreted in an analogous way. If we include these two residuals in the "accident equation"in the 2nd step as regressors, the regression coefficient of $\hat{\varepsilon}_i^1$ captures the effect of information asymmetry in the choice between no optional coverage ($Y_i = 0$) and some optional coverage ($Y_i = 1$) and the coefficient of $\hat{\varepsilon}_i^2$

[10]The exposition follows Kim et al. (2009), also the interpretation of the residuals below follows their presentation.

captures the effect of information asymmetry in the choice between some optional coverage ($Y_i = 1$) and all optional coverage ($Y_i = 2$).

Therefore in a second step we run a negative binomial regression or alternatively Poisson regression of the number of accidents Z_i in the contract year on the exogenous regressors including the generalized regressors calculated according to 5.6 and 5.7. The distribution of the number of accidents Z_i in the case of the negative binomial regression is given by

$$\mathbf{P}(Z_i) = \frac{\Gamma(Z_i + \frac{1}{\sigma^2}) \left[\sigma^2 \exp(X_i\hat{\beta}_0 + \hat{\varepsilon}_i\hat{\beta}_\varepsilon)\right]^{Z_i}}{\Gamma(\frac{1}{\sigma^2})\Gamma(Z_i+1) \left[1 + \sigma^2 \exp(X_i\hat{\beta}_0 + \hat{\varepsilon}_i\hat{\beta}_\varepsilon)\right]^{Z_i + \frac{1}{\sigma_i^2}}} \tag{5.8}$$

with Γ the Gamma function and $\hat{\beta}_0$, $\hat{\beta}_\varepsilon$ the estimated coefficient vectors. $\hat{\varepsilon}_i$ is defined as $(\hat{\varepsilon}_i^1, \hat{\varepsilon}_i^2)$.

Statistically significant and positive $\hat{\beta}_\varepsilon$ indicates the existence of asymmetric information between the parties.

As a two stage nonlinear estimation procedure is used one has to apply the Murphy-Topel standard error estimates in the second stage (Murphy and Topel (1985)). As in the second regression regressors are included which are estimated themselves in the first step one has to account for this additional source of uncertainty and correct the induced bias in the variance. An adapted version which is tailored to the situation above can be found in the appendix of Kim et al. (2009).

Allowing for nonlinearities (Dionne et al. (2001)) Dionne et al. (2001) choose the following procedure. In the first step $\hat{E}(Z|X)$ is computed by the estimation of a negative binomial regression of the distribution of accidents by using basic rating variables of the insurer as regressors. In the second step, a probit model with the chosen deductible as independent variable is run. The exogenous variables are the same as in the first step plus the expected number of claims estimated from the 1st step and the actual number of accidents.

In one of the first empirical studies, Puelz and Snow (1994) consider an ordered logit formulation for the deductible choice variable and find strong evidence for the presence of asymmetric information in the market for automobile collision insurance in Georgia. But Dionne et al. (2001) show that this correlation might be spurious because of the highly constrained form of the exogenous effects or the misspecification of the functional form used in the regression. They propose to add the estimate $\hat{E}(Z_i|X_i)$ of the conditional expected value of Z_i given X_i as a regressor into the ordered logit model to take into account the nonlinear effect of the risk classification variables, and by accounting for this, they find no residual asymmetric information in the market for Canadian automobile insurance.

2.5.2 Results

In this section we present our results. In order to check for robustness we apply all the parametric tests presented in the previous section to the Teilkasko and Vollkasko for both subsamples. This gives us a comprehensive picture and enables us to detect asymmetric information. The results for the whole portfolio are summarized in Table 2.7, the results for the novice drivers in Table 2.8. As the results for both groups are surprisingly relatively similar we will first present the results for the whole portfolio in some detail, especially the interpretation. Afterwards we will compare the results for the both subsamples.

In the Teilkasko the picture seems to be clear-cut. Both tests building up on the two Probits and the bivariate Probit reject the Null hypothesis of conditional independence on a significance level of $\alpha = 0.01$. In applying a 2 step estimation procedure the number of accidents has a significant influence in predicting the choice of deductible ($\hat{\beta}_{accidents} = 0.517$). In order to take into account nonlinearities, we also take up the expected number of accidents, additional to the number of accidents. The expected number of accidents is estimated according to a Poisson regression in which all variables used for risk classification by the insurance company are included. We also applied a Negative Binomial regression but as the results are similar to the Poisson regression and as we find no indication for overdispersion we omit them. While in Dionne et al. (2001) the addition of the expected number of accidents made the coefficient of the number of accidents insignificant our results are not changed so that the actual number of accidents remains a significant predictor for the choice of the deductible. Therefore also the two step estimation confirms the existence of asymmetric information in the Teilkasko. The multinomial approach shows that there is no asymmetric information in the choice between the highest possible deductible and some lower deductible. This information is contained in the coefficient of ε_1. In the choice between some deductible and full insurance, presumably the most important decision, the generalized residual ε_2 has a significant positive influence which indicates asymmetric information.

The interpretation of the results of the Vollkasko is similar but the results are contrary as can be seen in Table 2.7. The estimated coefficient of ρ is clearly not significantly different from zero and the test statistic W using the generalized residuals does not reject conditional independence on a significance level $\alpha = 0.01$. In the two step estimation the number of accidents is not a significant variable, regardless of taking into account nonlinearities. Also the multinomial approach shows no indication of asymmetric information if one examines the detailed choice between certain levels of deductible. The results for the young drivers are interpreted in an analogous way. A comparison of Table 2.7 and 2.8 shows that the pattern for the whole portfolio and the novice drivers is surprisingly similar and robust.

To sum up, we detect the existence of asymmetric information in the Teilkasko, but not in the Vollkasko, regardless whether the drivers are experienced or are novice drivers.

Table 2.7: Results for the whole portfolio (contracts signed after January 1, 2007)

Test Procedure	TK	VK
Two Probits	$W = 21.9$ reject conditional independence ($\alpha = 0.01, 0.05$)	$W = 0.01$ do not reject conditional independence ($\alpha = 0.01, 0.05$)
Bivariate Probit	$\rho = 0.286$ (0.049) reject conditional independence ($\alpha = 0.01, 0.05$)	$\rho = -0.1403$ (0.126) do not reject conditional independence ($\alpha = 0.01, 0.05$)
2 step estimation (Dionne et al. (2009))	$\hat{\beta}_{accidents} = 0.517^{*,**}$ (0.083)	$\hat{\beta}_{accidents} = 0.036$ (0.094)
	$\begin{cases} \hat{\beta}_{accidents} = 0.509^{*,**}(0.083) \\ \hat{\beta}_{exp\_acc} = 1.4(1.098) \end{cases}$	$\begin{cases} \hat{\beta}_{accidents} = 0.019(0.095) \\ \hat{\beta}_{exp\_acc} = 1.725(1.090) \end{cases}$
Multinomial approach	$\beta_{\hat{\varepsilon}_1} = 1.721$ (2.038/2.056)	$\beta_{\hat{\varepsilon}_1} = 0.338$ (0.332/0.332)
	$\beta_{\hat{\varepsilon}_2} = 0.352^{*,**}$ (0.079/0.149)	$\beta_{\hat{\varepsilon}_2} = 0.053$ (0.386/0.386)

Notes: *, ** indicate significance at the 1 % and 5 %, respectively.
Figures in brackets indicate the standard errors, and for the multinomial approach additionally the Murphy-Topel standard errors.

Table 2.8: Results for novice drivers

Test Procedure	TK	VK
Two Probits	$W = 15.4$ reject conditional independence ($\alpha = 0.01, 0.05$)	$W = 3.86$ do not reject conditional independence for $\alpha = 0.01$ and reject for $\alpha = 0.05$
Bivariate Probit	$\rho = 0.227$ (0.066) reject conditional independence ($\alpha = 0.01, 0.05$)	$\rho = -0.0101$ (0.108) do not reject conditional independence ($\alpha = 0.01, 0.05$)
2 step estimation (Dionne et al. (2009))	$\hat{\beta}_{accidents} = 0.388^{*,**}$ (0.075)	$\hat{\beta}_{accidents} = -0.123$ (0.076)
	$\begin{cases} \hat{\beta}_{accidents} = 0.386^{*,**}(0.076) \\ \hat{\beta}_{exp\_acc} = 0.640(1.375) \end{cases}$	$\begin{cases} \hat{\beta}_{accidents} = -0.125(0.076) \\ \hat{\beta}_{exp\_acc} = 1.306(1.871) \end{cases}$
Multinomial approach	$\beta_{\hat{\varepsilon}_1} = 0.1241$ (0.2676)	$\beta_{\hat{\varepsilon}_1} = 0.038$ (0.167)
	$\beta_{\hat{\varepsilon}_2} = 0.4136^{*,**}$ (0.074)	$\beta_{\hat{\varepsilon}_2} = -1.785$ (4.086)

Notes: *, ** indicate significance at the 1 % and 5 %, respectively.
Figures in brackets indicate the standard errors, and for the multinomial approach additionally the Murphy-Topel standard errors.

2.6 The Generalized Positive Correlation Property

2.6.1 Statistical Procedures

As mentioned earlier, Chiappori et al. (2006) extend the positive correlation property (for suitably defined notions) to much more general models, entailing heterogeneous preferences, multiple level of losses, multidimensional adverse selection plus possibly moral hazard and even nonexpected utility. Their generalized positive correlation property relies on several general assumptions and generalizes the notion of positive correlation between risk and coverage. The most important assumptions among these are the assumption of "realistic expectations"and the "nonincreasing profit"(NIP) assumption. "Realistic expectations" means that when agents choose a contract, they correctly assess their accident probability and loss distribution. With other words, they know their loss distribution. The NIP assumption states that if a contract C_2 covers more than a contract C_1, then the expected profits generated by C_2 are not higher than profits under C_1: $\pi(C_2) \leq \pi(C_1)$.
These assumptions resp. the implied predictions can be tested with the data and the results

are of interest on their own. As we apply them to contracts with deductibles we present only the special cases for contracts with deductibles. In the following C_1 and C_2 are contracts with straight deductibles, where C_2 covers more than C_1, i.e., the deductible for contract 1 is higher, $d_1 \geq d_2$. P_1, P_2 denote the corresponding premia, R_1, R_2 denote the indemnities for every possible claim under each contract and can be approximated by loss minus the deductible in our case.

Under the null that the insurees know their loss distribution, i.e., have realistic expectations, the following property should hold in the case of straight deductibles

$$P_2 - P_1 \geq q_1(d_1 - d_2)$$

q_1 denotes the probability that the loss L is above the deductible d_1 under C_1: $q_1 = Prob(L > d_1)$.

The condition of NIP can be formulated in this special case as

$$P_1/(1+t) - P_2/(1+t) \geq (1+\lambda)(R_1 - R_2)$$

with t the tax rate and λ the loading factor.

Finally, under the two assumptions mentioned above (and some other weak conditions) a generalized positive correlation property holds which in our case can be written as[11]

$$(1+K)(E_2 - E_1) \geq d_2(q_2 - q_1).$$

K is defined as $(1+t)(1+\lambda) - 1$, E_i denotes the expected loss under contract i. This can be approximated by the payment of compensation in the case of an accident plus the corresponding deductible. q_i is the probability of a claim under contract C_i.

It is important to mention that the inequalities above are only valid for individuals with the same observables X and therefore have to be checked for all cells separately.

In order to test if these inequalities hold one forms a test statistic T for each cell. In the case of realistic expectations T is given by

$$T = P_2 - P_1 - q_1(d_1 - d2)$$

for the NIP

$$T = P_1/(1+t) - P_2/(1+t) - (1+\lambda)(R_1 - R_2)$$

[11] Chiappori et al. (2006) give further explanations and an intuition for the result.

and finally for the positive correlation property

$$T = (1 + K)(E_2 - E_1) - d_2(q_2 - q_1).$$

As these test statistics are calculated for every cell we get an empirical distribution for each test statistic. Under the null hypothesis that $T(X) = 0$ for all X, these numbers should be distributed as a standard normal distribution $N(0,1)$. There are several ways to come to a test decision. One way is to calculate the standardized mean and to conduct a t-test if the mean is different from zero. Additionally we report the share of the number of positive signs of our test statistic as we think that this offers valuable additional information to get an overview over the dispersion of the distribution. All variables which are required are in the data set or can be estimated. As we have in the TK and VK several levels of deductibles we compare them pairwise.

2.6.2 Results

In this section we present the results of the above tests according to TK and VK. We applied the tests to the whole portfolio of contracts in the year 2009. In order to save space we only present the results if one of the deductibles involved in the comparisons contains at least more than 5 % of the contracts of the total portfolio.

The first question to address is if the customers have realistic expectations. Our results show that for both the TK and VK and for the different choices of deductibles the null can be rejected on any reasonable significance level and therefore the customers have realistic expectations about their true loss distribution.

Table 2.9: Realistic expectations - VK 2009

coverage 1	coverage 2	number of cells	std. mean	t-statistic	share of pos. signs
300	150	6,843	1.90	157.55	0.998
500	150	6,595	2.19	178.23	1.000
500	300	8,730	1.89	176.43	0.999

coverage in Euro. sample size appr. 2.3 mio.

Table 2.10: Realistic expectations - TK 2009

coverage 1	coverage 2	number of cells	std. mean	t-statistic	share of pos. signs
150	0	1,880	1.72	74.68	0.999

coverage in Euro. sample size appr. 1.2 mio.

The second question to be addressed is if the non increasing profit assumption holds. This is also interesting with regard to the market structure resp. competition on this market. Tables 2.11 and 2.12 below show that the mean is significantly different from zero and therefore the null can be rejected. But the share of observations with positive sign falls below 50 %. Therefore the situation is not completely clear as the value of the test statistic suggests, but there is a tendency that the NIP assumption is fulfilled.

Table 2.11: Non increasing profit assumption (NIP) - VK 2009

coverage 1	coverage 2	number of cells	std. mean	t-statistic	share of pos. signs
300	150	6,843	0.04	3.70	0.38
500	150	6,595	0.07	5.56	0.44
500	300	8,730	0.02	2.09	0.62

coverage in Euro. sample size appr. 2.3 mio.

Table 2.12: Non increasing profit assumption (NIP)- TK 2009

coverage 1	coverage 2	number of cells	std. mean	t-statistic	share of pos. signs
150	0	1,880	0.06	2.57	0.450

coverage in Euro. sample size appr. 1.2 mio.

Finally we present the results of the general positive correlation tests, summarized in Tables 2.13 and 2.14. For the Teilkasko we can reject the null hypothesis (on a significance level of 1 %) and therefore confirm the result of asymmetric information in this part of the automobile insurance. The share of the cells with positive sign is approximately 75.6 %. This result is in line with the results we received by applying the parametric procedures of the previous section. For the VK the picture is not so clear cut. The tests confirm the existence of a positive correlation property. This contradicts the previous section at a first glance. But the number of cells with positive sign is quite low. For the comparisons of deductibles which we omitted the share goes down to about one fourth. Therefore the overwhelming number of cells has negative sign. This makes it plausible that despite of our test statistic the positive correlation property does not hold for most cells and therefore the existence of asymmetric information seems to be doubtful resp. the effect is not strong. Moreover, a detailed look at the cells reveals that the high value of the test statistics is driven by some cells with a sparse number of observations which could be classified as outliers.

Summing up, we can confirm that the drivers have realistic expectations with regard to their loss distribution and that in this market (and this period) the NIP condition holds. Moreover, we show that the generalized positive correlation property holds for the Teilkasko. For the Vollkasko the picture is not so clear cut. Depending on the chosen test, evidence for the Vollkasko that this property also holds is quite weak.

Table 2.13: Generalized positive correlation - VK 2009

coverage 1	coverage 2	number of cells	std. mean	t-statistic	share of pos. signs
300	150	6,843	0.09	7.56	0.410
500	150	6,595	0.15	12.26	0.495
500	300	8,730	0.08	7.45	0.659

coverage in Euro. sample size appr. 2.3 mio.

Table 2.14: Generalized positive correlation - TK 2009

coverage 1	coverage 2	number of cells	std. mean	t-statistic	share of pos. signs
150	0	1,880	0.40	17.42	0.756

coverage in Euro. sample size appr. 1.2 mio.

2.7 Discussion and Concluding Remarks

In this paper, we analyzed a data set on automobile insurance in Germany. This market is of interest not only because it is the largest such market in Europe, but also because particular contractual arrangements allow us to analyze asymmetric information with respect to different kind of risks which are usually covered by car insurance.

Our first conclusion is that in Teilkasko (partial insurance), asymmetric information exists. This is confirmed by the tests in section 5 and 6. By weighing all test results we would deny the existence of asymmetric information in the Vollkasko resp. the effect is very weak compared to the Teilkasko. This is valid for both the young driver and the whole portfolio.

The Vollkasko covers damages to the own car or own body which are not covered in the Teilkasko. This are especially damages in the case of an accident for which one is at his own fault. Damages to the other party are covered by the Haftpflichtversicherung. In contrast to the Teilkasko there is a bonus / malus coefficient in the Vollkasko, i.e., an accident for which the policy holder is liable leads to a worsening of this coefficient and thus to a higher premium in the following years. As there is an incentive in the Vollkasko not to file all accidents we limited our analysis to claims which are twice as high as the highest deductible, i.e., we considered only damages above $2,000$ Euro. Accidents with this magnitude are in nearly almost cases filed as a claim to the insurance company. In the Vollkasko we find no convincing evidence for the existence of asymmetric information.

When we restrict our analysis to "young driver", i.e., drivers with no driving experience it is reasonable to assume that these insured persons do not have an informational advantage concerning their accident probability. This means that in the Vollkasko case we can exclude adverse selection and therefore the analysis can be interpreted as a test for moral hazard. But our results show that there is no positive correlation as predicted by the theory of moral hazard. This might be due to the existence of the experience rating which is in a way deterrent and prohibitive and does not exist in the Teilkasko.

If one does not want to maintain the assumption that young drivers have no informational advantage than different effects must cancel each other which might be possible but not very probable.

If we have a look at the whole portfolio the pattern is the same. Drivers have an informational advantage in the Teilkasko. The absence of asymmetric information in the Vollkasko might be explained in the following way: the bonus / malus coefficient is on the one side a good proxy for the ability and driving history of the driver and on the other side a suitable scheme to induce careful driving.

One could presume that in the Teilkasko in particular risks are covered which cannot be influenced by individual behavior to the same extent as risks in the Vollkasko, so that the influence of individual behavior is not the most important factor. Therefore one might argue that moral hazard is of minor importance in the Teilkasko so that the positive correlation is driven by adverse selection. This means that the insurees are aware of potential risks like storms, etc. and choose the level of deductible (full insurance / deductible) according to their risk exposure.

Another result is that the insurees know their loss distribution, i.e., have realistic expectations. This rules out the possible explanation that insurees have different abilities in assessing the risks related to the Teilkasko or Vollkasko. We can also conclude that drivers do not over- or underestimate their risk as it is often conjectured.

We also checked the NIP assumption and found that its holds. Usually insurance markets are characterized by an oligopolistic market structure, but this results confirms our conjecture from section 3 that the German market during this period is close to perfect competition.

2.8 References

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

Bloomberg (11/13/2009). Allianz wont cut auto premiums, braced to lose market share. published in `www.bloomberg.com`.

Chiappori, P.-A., Heckman, J. J., 1999. Testing for moral hazard using dynamic data. Manuscript, Dept. of Economics, Univ. of Chicago.

Chiappori, P.-A., Salanié, B. 1997. Empirical contract theory: the case of insurance data. European Economic Review 41, 943-950.

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

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., 2005. Asymmetric information and learning: evidence from the automobile insurance market. The Review of Economics and Statistics 87, 197-207.

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

Crocker, K.-J., Snow, A., 2000. The theory of risk classification. In: G. Dionne, editor, Handbook of Insurance, Chapter 8, 245-276. Kluwer Academic Publishers.

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

Dionne, G., Gourieroux C., 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., Michaud, P.-C., Dahchour, M., 2004. Separating moral hazard from adverse selection in automobile insurance: Longitudinal evidence from France. Cahiers de recherche 0420, CIRPEE.

GenRe (2010). Die Kfz-Versicherung im Jahr 2010: Ende des Preiskampfs oder Preiskampf ohne Ende?.

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.

Handelsblatt (09/01/2010). Allianz opfert die Marktführerschaft in der Auto-Versicherung. published in `www.handelsblatt.com`.

Jullien, B., Salanié, B., Salanié, F., 2007. Screening risk averse agents under moral hazard. Economic Theory 30, 151-169.

Kim, H., Kim, D., Im, S., Hardin, J. W., 2009. Evidence of asymmetric information in the automobile insurance market: dichotomous versus multinomial measurement of insurance coverage. The Journal of Risk and Insurance 76, 343-366.

Muermann, A., Straka, D., 2011. Asymmetric information in automobile insurance: new evidence from telematics data. Working paper.

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

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

Reuters (01/28/2009). German motor insurance competition still hard - HUK. published in `www.reuters.com`.

Reuters (05/17/2010). German car insurance prices up, market murky - HUK. published in `www.reuters.com`.

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.

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.

Saito, K., 2009. Testing for asymmetric information in the automobile insurance market under rate regulation. The Journal of Risk and Insurance 73, 335-356.

Spindler, M., Su, L., 2011. Nonparametric testing for asymmetric information. Working Paper.

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

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`.

Zavadil, T., 2011. Do the better insured cause more damage? Testing for asymmetric information in car insurance. Working Paper.

Chapter 3

Nonparametric Testing for Asymmetric Information*

3.1 Introduction

Since Akerlof (1970) the notion of asymmetric information, comprising adverse selection and moral hazard, has been explored at a rapid pace. At the same time people observed a wide gap between the theoretical development and empirical studies in asymmetric information. This gap has recently become narrower. In particular, the insurance market has been a fruitful and productive field for empirical studies. There are two reasons for this. First, insurance contracts are usually highly standardized and can completely be described by a relatively small set of variables, and insurees' performances, i.e., the occurrence of a claim and possibly its cost, are exactly filed in the database of an insurance company. Second, insurance companies have hundreds of thousands or even millions of clients and therefore the samples are sufficiently large for econometric studies. Hence, fields like automobile insurance, annuities and life insurance, crops insurance, long-term care and health insurance offer a large sample of standardized contracts for which performances are recorded and therefore are well suited for testing the theoretical predictions of contract theory. For a detailed justification for using insurance data to test contract theory, see Chiappori and Salanié (1997). For a recent overview over the issue of testing for adverse selection in insurance markets, see Cohen and Siegelman (2010). The latter paper covers a large number of empirical studies in different insurance branches.

In statistical terms, the theoretical notion of asymmetric information implies a positive (conditional) correlation between coverage and risk as both adverse selection and moral haz-

*This chapter is based on joint work with Liangjun Su.

ard predict this positive correlation. In their seminal paper Chiappori and Salanié (2000) propose both parametric and nonparametric methods to test this. Their nonparametric tests are restricted to discrete data with only two categories per variable even though some of the variables in the data set are continuous and others have far more than two categories. Therefore, in order to conduct Chiappori and Salanié's nonparametric test, all variables must be transformed to binary variables, which often results in a loss of information. Following the lead of Chiappori and Salanié (2000), most subsequent studies use a variation of their parametric testing procedure which has become somewhat standard in the empirical contract theory. Nevertheless, these parametric tests are fragile to both *functional* and *distributional* form misspecifications which are a severe problem in this field. For example, in the automobile insurance market it is common knowledge that the age of the driver has a nonlinear effect on the probability of an accident, but such a nonlinear effect has rarely been taken into account in the literature. For another example, the error term in the binary model for modeling the choice of an insurance contract may not be either normally or logistically distributed, and tests for asymmetric information based on the probit or logit model can therefore yield misleading conclusions in the case of incorrect distributional specification. For this reason, in this paper we propose a new purely nonparametric test for asymmetric information based on the notion of conditional independence, which avoids the problem of either functional or distributional misspecification.

The absence of asymmetric information means that the choice of a contract Y (discrete variable) provides no information for predicting the "performance" variable Z (discrete or continuous, e.g., the number of claims or the sum of reimbursements), conditional on the vector X of all exogenous variables (discrete and continuous). Therefore we can transform the problem of testing the absence of asymmetric information into a test for conditional independence: $F(Z|X,Y) = F(Z|X)$ almost surely (a.s.) where, e.g., $F(Z|X,Y)$ denotes the conditional cumulative distribution function (CDF) of Z given (X,Y). We propose a nonparametric test statistic to test the conditional independence of Z and Y given X. We show that the test statistic is asymptotic normally distributed under the null hypothesis of conditional independence (or absence of asymmetric information) and diverges to infinity in the presence of conditional dependence (or asymmetric information). We then apply our test to a French automobile insurance data set and compare our testing results with the results found in the literature.

The rest of the paper is structured as follows. Section 2 outlines the theory of asymmetric information. Section 3 reviews the standard statistical tools for testing asymmetric information. We introduce a new nonparametric test for conditional independence in Section 4.

We conduct a small set of Monte Carlo simulations to examine the performance of the new test in Section 5. We apply our test to test for the asymmetric information in the French insurance market in Section 6. Final remarks are contained in Section 7. All technical details are relegated to the Appendix.

3.2 The Theory of Asymmetric Information

In their seminal paper Rothschild and Stiglitz (1976) introduce the notion of adverse selection in insurance markets that has been extended in many directions since then.[1] In the basic model, the insureds have private information about the expected claim, exactly speaking about the probability that a claim with fixed level occurs, while the insurers do not have this information. Thus there are two groups with different claim probabilities, the "bad" and "good" risks. The agents have identical preferences which are moreover perfectly known to the insurer. Additionally, perfect competition and exclusive contracts are assumed. Exclusive contracts mean that an insuree can buy coverage only from one insurance company. This allows firms to implement nonlinear (especially convex) pricing schemes which are typical under asymmetric information. Under this setting insurance companies offer a menu of contracts in equilibrium: a full insurance which is chosen by the "bad" risks and a partial coverage which is bought by the "good" risks. In general, contracts with more comprehensive coverage are sold at a higher (unitary) premium.

Therefore, one expects a positive correlation between "risk" and "coverage" (conditional on observables). Since the assumptions in the Rothschild and Stiglitz model are very simplistic and normally not fulfilled in real applications, an important question to address is how robust this coverage-risk correlation is. Chiappori et al. (2006) show that the positive correlation property extends to much more general models, as already conjectured by Chiappori and Salanié (2000). Under competitive markets this property is also valid in a very general framework entailing heterogeneous preferences, multiple level of losses, multidimensional adverse selection plus possible moral hazard and even non-expected utility theory. In the case of imperfect competition some form of positive correlation holds if the agent's risk aversion is public information. In the case of private information the property does not necessarily hold (c.f. Jullien et al. (2007)).

While adverse selection concerns "hidden information", moral hazard deals with "hidden action". Moral hazard occurs when the expected loss (accident probability or level of damage)

[1] For a detailed survey on adverse selection and the related moral hazard problem, see Dionne, Doherty and Fombaron (2000) and Winter (2000), respectively.

is not exogenous, as assumed in the adverse selection case, but depends on some decision or action made by the subscriber (e.g., effort or prevention) which is neither observable nor contractible. A higher coverage leads to decreased efforts and therefore to a higher expected loss. Therefore moral hazard also predicts a positive correlation between "coverage" and "risk".

Although both phenomena lead to a positive risk-coverage correlation, there is one important difference: under adverse selection the risk of the potential insuree affects the choice of the contract, whereas under moral hazard the chosen contract influences the behavior and therefore the expected loss. So there exists reversed causality in both cases.[2]

In sum, the theory of asymmetric information predicts a positive correlation between (appropriately defined) "risk" and "coverage" which should be quite robust.[3]

To proceed, it is worth mentioning that to test for asymmetric information, the researcher needs access to the same information which is also available to the insurer and used for pricing. The theory of adverse selection predicts that the insurance company offers a menu of contracts to indistinguishable individuals. Individuals are (ex ante) indistinguishable for the insurer if they share the same characteristics. Therefore the positive risk-coverage correlation is valid only conditional on the observed characteristics. Different groups of observable equivalent individuals are offered different menus of contracts with different prices according to their risk exposure.[4] Only within each class are the mechanisms described above valid.

3.3 Standard Testing Procedures

In this section we review some tests of asymmetric information in the literature. We first outline the general structure of the problem and then review the parametric and nonparametric testing procedures in turn.

3.3.1 General Structure

In the following we denote by X the vector of exogenous control variables to be conditional on, by Y a decision or choice variable, and by Z the endogenous "performance" variable. In the context of insurance, X usually includes variables that are used for risk classification by

[2] To disentangle moral hazard from adverse selection is an important problem in the empirical literature. The first attempt is Dionne et. al. (2004). An overview over different possible strategies for dealing with this problem can be found in Cohen and Siegelman (2010).

[3] It seems that in the empirical insurance literature adverse selection is more stressed than the moral hazard aspect which only receives minor attention, see, e.g., Cohen and Siegelman (2010).

[4] For the theory of risk classification under asymmetric information see Crocker and Snow (2000).

the insurance company, Y could be the choice of deductibles, and Z could be the number of accidents or claims or the sum of reimbursements caused by accidents.[5] As we shall see, we allow both continuous and discrete variables in X, and Z can be continuous or discrete. For concreteness, we assume that Y is a discrete variable. There is no asymmetric information if and only if the prediction of the endogenous variable Z based on X and Y jointly coincides with its prediction based on X alone. Formally, this can be stated in terms of the equivalence of two conditional CDFs:

$$F(Z|X,Y) = F(Z|X) \text{ a.s.,} \tag{3.1}$$

where, e.g., $F(Z|X,Y)$ denotes the conditional CDF of Z given (X,Y). Intuitively, this means that the choice of the contract, e.g., the choice of a certain deductible, provides no useful information for predicting the risk, e.g., the number of claims, as soon as the risk classes are controlled for. Equivalently, we can interchange the roles of Z and Y :

$$F(Y|X,Z) = F(Y|X) \text{ a.s.,} \tag{3.2}$$

where, e.g., $F(Y|X,Z)$ denotes the conditional CDF of Y given (X,Z). (3.2) says that the number of claims (or the sum of reimbursements caused by accidents) does not provide useful information to predict the choice of deductibles as long as we control the risk classes. Either (3.1) or (3.2) indicates the conditional independence of Y and Z given X.[6]

3.3.2 Parametric Testing Procedures

Almost all empirical studies analyzing the positive risk-coverage correlation property use one of the following two types of parametric procedures.

The first approach is to run a regression of Z_i on Y_i and X_i and to test whether the coefficient of Y_i is zero or not. When Z_i is continuously valued, the regression model is

$$Z_i = \beta_0 + \beta_1 Y_i + \beta_2' X_i + \varepsilon_i, \tag{3.3}$$

where ε_i is the error term, β_0, and $\left(\beta_1, \beta_2'\right)$ are intercept and slope coefficients, respectively,

[5] The distinction of accidents and claims is a very important point in the empirical literature as not every accident leads to a claim. Neglecting this issue might lead to biased results.

[6] Alternatively, one can use conditional probability density or mass functions to form the independence between Y and Z conditional on X : $f(Z|X,Y) = f(Z|X)$, or $f(Y|X,Z) = f(Y|X)$ a.s., where, e.g., $f(Z|X,Y)$ denotes the conditional probability density or mass function of Z given (X,Y). See Su and White (2007, 2008, 2010) for other equivalent formulations.

and the prime denotes transpose. When Z_i is a dummy variable, the regression model is

$$Z_i = \mathbf{1}(\beta_0 + \beta_1 Y_i + \beta_2' X_i + \varepsilon_i > 0), \tag{3.4}$$

where ε_i is assumed to be either normally or logistically distributed, and $\mathbf{1}(A) = 1$ if A is true and 0 otherwise. If Z_i is a discrete variable that has more than two categories, then one can use the ordered logit model. One obvious drawback of this approach is that it neglects by construction the potential nonlinear effects of the controlled variables, and a test based on (3.3) is designed to test the conditional mean independence of Z_i and Y_i given X_i, which is a much weaker condition than conditional independence at the distributional level. In addition, the distributional assumption in the probit, logit, or ordered logit model may not hold, and once this happens, tests for asymmetric information can lead to misleading conclusions.

In one of the first empirical studies Puelz and Snow (1994) consider an ordered logit formulation for the deductible choice variable and find strong evidence for the presence of asymmetric information in the market for automobile collision insurance in Georgia. But Dionne et al. (2001) show that this correlation might be spurious because of the highly constrained form of the exogenous effects or the misspecification of the functional form used in the regression. They propose to add the estimate $\hat{E}(Z_i|X_i)$ of the conditional expected value of Z_i given X_i as a regressor into the ordered logit model to take into account the nonlinear effect of the risk classification variables, and by accounting for this, they find no residual asymmetric information in the market for Canadian automobile insurance.

A second and more advanced approach was introduced by Chiappori and Salanié (1997, 2000) and has become widespread in the empirical contract theory since then. They define two probit models, one for the choice of the coverage Y_i (either compulsory/basic coverage or comprehensive coverage) and the other for the occurrence of an accident Z_i (either no accident being blamed for or at least one accident with fault):

$$\begin{cases} Y_i = \mathbf{1}(\beta' X_i + \varepsilon_i > 0) \\ Z_i = \mathbf{1}(\gamma' X_i + \eta_i > 0) \end{cases} \tag{3.5}$$

where ε_i and η_i are independent standard normal errors, and β and γ are coefficients. They first estimate these two probit models independently, calculate the generalized residuals $\hat{\varepsilon}_i$

and $\hat{\eta}_i$, and then construct the following test statistic[7]

$$W_n = \frac{\left(\sum_{i=1}^n \hat{\varepsilon}_i \hat{\eta}_i\right)^2}{\sum_{i=1}^n \hat{\varepsilon}_i^2 \hat{\eta}_i^2}. \tag{3.6}$$

Under the null of conditional independence, $\text{cov}(\varepsilon_i, \eta_i) = 0$ and W_n is distributed asymptotically as $\chi^2(1)$. Alternatively, one can estimate a bivariate probit model in which ε_i and η_i are distributed as bivariate normal with correlation coefficient ρ to be estimated, and then test whether $\rho = 0$ or not. They find no evidence of asymmetric information in the French automobile insurance market.

3.3.3 Nonparametric Testing Procedures

Motivated by the χ^2-test for independence in the statistics literature, Chiappori and Salanié (2000) propose a nonparametric test for asymmetric information by restricting all variables in X_i, Y_i, and Z_i to be binary. They choose a set of m exogenous binary variables in X_i, and construct $M \equiv 2^m$ cells in which all individuals have the same values for all variables in X_i. For each cell they set up a 2×2 contingency table generated by the binary values of Y_i and Z_i, and conduct a χ^2-test for independence. This results in M test statistics, each of which is distributed asymptotically as $\chi^2(1)$ under the null hypothesis. They aggregate these M test statistics in three ways to obtain three overall test statistics for conditional independence: one is the Kolmogorov-Smirnoff test statistic that compares the empirical distribution function of the M test statistics with the CDF of the $\chi^2(1)$ distribution; the second is to count the number of rejections for the independence test for each cell which is asymptotically distributed as binomial $B(M, \alpha)$ under the null, where α denotes the significance level of the χ^2 test within each cell; and the third is the sum of all the test statistics for each individual cell, which is asymptotically $\chi^2(M)$ distributed under the null. Again, using these nonparametric methods, they find no evidence for the presence of asymmetric information in the French automobile insurance market.

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

3.4 A New Nonparametric Test

In this section we propose a new nonparametric test for asymmetric information based on the formulation in (3.1). The null hypothesis is

$$H_0 : F(Z|X,Y) = F(Z|X) \text{ a.s.,} \tag{4.1}$$

and the alternative hypothesis is

$$H_1 : \Pr\{F(Z|X,Y) = F(Z|X)\} < 1. \tag{4.2}$$

We consider the case where Y is a discrete random variable (typically a dummy variable), Z can be either discrete or continuous, and X contains both continuous and discrete variables. Note that early literature on testing for conditional independence mainly focus on the case where both Y and X are continuously distributed, see, Delgado and González-Manteiga (2001), Su and White (2007, 2008, 2010), Song (2009), Huang (2009), Huang and White (2009), to name just a few. Even though we restrict our attention mainly on the case where Y is discrete, we remark that in the case of continuous Y, the proposed test continues to work with little modification.

3.4.1 The Test Statistic

Given observations $\{(X_i, Y_i, Z_i)\}_{i=1}^n$, one could propose a test based on the comparison of two conditional cumulative distribution (CDF) estimates, one is the conditional CDF of Z given X $(F(z|x))$ and the other is the conditional CDF of Z given (X,Y) $(F(z|x,y))$.Nevertheless, for the reason elaborated at the end of this section, we will compare $F(z|x,y)$ with $F(z|x,\widetilde{y})$ for different values y and $\widetilde{y}$ instead.

For more rigorous notation, one could use $F_{Z|X}(z|x)$ $(F_{Z|X,Y}(z|x,y))$ to denote the conditional CDF of Z given X $((X,Y))$. Below we make reference to these CDFs and several probability density functions (PDFs) simply using the list of their arguments – for example, $p(x,y,z)$, $p(x,y)$ and $p(x)$ denote the PDFs of (X_i, Y_i, Z_i), (X_i, Y_i), and X_i, respectively. This notation is compact, and we hope, sufficiently unambiguous. In addition, even though a PDF is most commonly associated with continuous distributions, here we use it to denote the Radon–Nikodym derivative of a CDF with respect to the Lebesgue measure for the continuous component and the counting measure for the discrete component.

To allow for both continuous and discrete regressors in X_i, write $X_i = (X_i^{c'}, X_i^{d'})'$ where

X_i^c denotes a $p_c \times 1$ vector of continuous regressors in X_i and X_i^d denotes a $p_d \times 1$ vector of remaining discrete regressors with $p_d \equiv p - p_c$. For simplicity, we assume that none of the discrete regressors has a natural ordering and each takes only a finite number of values.[8] We use X_{is}^c (X_{is}^d) to denote the sth component of X_i^c (X_i^d), where $s = 1, \cdots, p_c$ (p_d). We assume that X_{is}^d takes c_s different values in $\mathcal{X}_s^d \equiv \{0, 1, \cdots, c_s - 1\}$, $s = 1, \cdots, p_d$, and Y_i takes c_y different values in $\mathcal{Y} \equiv \{0, 1, \cdots, c_y - 1\}$.

Fix $y \in \mathcal{Y}$. We consider the estimation of $F(z|x, y)$ by using the local linear method. For this purpose, we define the kernels for the continuous regressor X_i^c and discrete regressor X_i^d separately. For the continuous regressor, we choose a product kernel function $Q(\cdot)$ of $q(\cdot)$ and a vector of smoothing parameters $h \equiv (h_1, ..., h_{p_c})'$. Let $Q_{h,j}(x^c) \equiv \Pi_{s=1}^{p_c} h_s^{-1} q\left(\left(X_{js}^c - x_s^c\right)/h_s\right)$ and

$$Q_{h,ji} \equiv Q_h\left(X_j^c - X_i^c\right) = \prod_{s=1}^{p_c} h_s^{-1} q\left(\left(X_{js}^c - X_{is}^c\right)/h_s\right), \tag{4.3}$$

where, for example, $x^c \equiv (x_1^c, \cdots, x_{p_c}^c)'$, and X_{is}^c denote the sth element in X_i^c. For the discrete regressor, we follow Racine and Li (2004) and Li and Racine (2007, 2008) and use a variation of the kernel function of Aitchison and Aitken (1976):

$$l\left(X_{js}^d, X_{is}^d, \lambda_s\right) = \begin{cases} 1 & \text{if } X_{js}^d = X_{is}^d \\ \lambda_s & \text{otherwise} \end{cases} \tag{4.4}$$

where $\lambda_s \in [0, 1]$ is the smoothing parameter. In the special case where $\lambda_s = 0$, $l(\cdot, \cdot, \cdot)$ reduces to the usual indicator function as used in the nonparametric frequency approach. Similarly, $\lambda_s = 1$ leads to a uniform weight function, in which case, the X_{is}^d regressor will be completely smoothed out in the sense that it will not affect the nonparametric estimation result. The product kernel function for all the discrete vectors is given by

$$L_{\lambda,ji} \equiv L_\lambda\left(X_j^d, X_i^d\right) \equiv \prod_{s=1}^{p_d} \lambda_s^{\mathbf{1}(X_{js}^d \neq X_{is}^d)}, \tag{4.5}$$

where $\lambda \equiv (\lambda_1, \cdots, \lambda_{p_d})'$. Combining (4.3) and (4.5), we obtain the product kernel function for the conditioning vector X_i:

$$K_{h\lambda,ji} \equiv K_{h\lambda}\left(X_j, X_i\right) = Q_h\left(X_j^c - X_i^c\right) L_\lambda\left(X_j^d, X_i^d\right). \tag{4.6}$$

[8]When some of the conditioning variables in X_i have a natural ordering, one can easily modify the discrete kernel defined below following either Racine and Li (2004) or Li and Racine (2007, 2008).

Now, fix a point $X_i = (X_i^{c\prime}, X_i^{d\prime})'$. It follows from the first order Taylor expansion that

$$F(z|X_j, y) \approx F(z|X_i, y) + \dot{F}(z|X_i, y)' \left(X_j^c - X_i^c\right) \tag{4.7}$$

for any X_j^c in the neighborhood of X_i^c and $X_j^d = X_i^d$, where $\dot{F}(z|x, y) = \partial F(z|x)/\partial x^c$, i.e., the derivative is only taken with respect to the continuous component x^c of $x \equiv (x^{c\prime}, x^{d\prime})'$. Given observations $\{(X_i, Y_i, Z_i)\}_{i=1}^n$, we estimate $F(Z_i|X_i, y)$ by solving the weighted least squares minimization problem

$$\min_{\boldsymbol{\beta}} \sum_{j=1}^{n} \left[\mathbf{1}\{Z_j \leq Z_i\} - \beta_0 - \beta_1' \left(\left(X_j^c - X_i^c\right)/h\right)\right]^2 K_{h\lambda, ji} \mathbf{1}_j^y, \tag{4.8}$$

where $\boldsymbol{\beta} \equiv \left(\beta_0, \beta_1'\right)'$ and $\mathbf{1}_j^y \equiv \mathbf{1}(Y_j = y)$. Our estimator $\widehat{F}(Z_i|X_i, y)$ is the minimizing intercept term in the above problem. Let $\boldsymbol{\tau}_h\left(X_j^c - x^c\right) \equiv \left(1, \left(\left(X_j^c - x^c\right)/h\right)'\right)'$. Then it is easy to verify that

$$\widehat{F}(Z_i|X_i, y) = e_1' \left[\mathbf{S}_{ny}(X_i)\right]^{-1} \frac{1}{n} \sum_{j=1}^{n} K_{h\lambda, ji} \mathbf{1}_j^y \boldsymbol{\tau}_h\left(X_j^c - X_i^c\right) \mathbf{1}(Z_j \leq Z_i)$$

where $e_1 \equiv (1, 0, \cdots, 0)'$ is a $(p_c + 1)$-vector, and
$\mathbf{S}_{ny}(X_i) \equiv \frac{1}{n} \sum_{j=1}^{n} K_{h\lambda, ji} \mathbf{1}_j^y \boldsymbol{\tau}_h(X_j^c - X_i^c) \boldsymbol{\tau}_h(X_j^c - X_i^c)'$.

We measure the variations in $\widehat{F}(Z_i|X_i, y)$ across different values of y and different observations by

$$D_n \equiv \sum_{r=0}^{c_y - 2} \sum_{s=r+1}^{c_y - 1} \sum_{i=1}^{n} \left[\widehat{F}(Z_i|X_i, r) - \widehat{F}(Z_i|X_i, s)\right]^2.$$

We study the asymptotic properties of D_n under H_0, a sequence of Pitman local alternatives, and the global alternative H_1. We will show that after being appropriately recentered and scaled, D_n is asymptotically normally distributed under the null and local alternatives, and diverges to infinity under the global alternative.

3.4.2 Assumptions

Throughout the paper we use ξ_i, ζ_i, and ς_i to denote $(X_i', Y_i, Z_i)'$, $(X_i', Y_i)'$, and $(X_i', Z_i)'$, respectively. Similarly, let $\xi \equiv (x', y, z)'$, $\zeta \equiv (x', y)'$ and $\varsigma \equiv (x', z)'$. With a little bit abuse of notation, we use $p(\xi)$, $p(\zeta)$, and $p(x)$ to denote the PDF of ξ_i, ζ_i, and X_i, respectively. Similarly, $F(z|x, y) \equiv F(z|x^c, x^d, y)$ denotes the conditional CDF of Z_i given $(X_i^{c\prime}, X_i^{d\prime}, Y_i)'$.

To facilitate our asymptotic analysis, we make the following assumptions.

Assumption A.1 The sequence $\{\xi_i\}_{i=1}^{n}$ is independent and identically distributed (IID) with CDF F_ξ.

Assumption A.2 (i) The support $\mathcal{X}^c$ of X_i^c is compact.

(ii) $p(\xi)$ is uniformly bounded over its support $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, where $\mathcal{X} \equiv \mathcal{X}^c \times \mathcal{X}^d$, $\mathcal{X}^d \equiv \mathcal{X}_1^d \times \cdots \times \mathcal{X}_{p_d}^d$, and $\mathcal{Z}$ is the support of Z_i. $p(\zeta) \equiv p\left(x^c, x^d, y\right)$ is bounded away from 0 for all $x^c \in \mathcal{X}^c$, $x^d \in \mathcal{X}^d$, and $y \in \mathcal{Y}$.

Assumption A.3 Let $\eta \equiv \left(x^d, y\right)$. (i) For each $\eta \in \mathcal{X}^d \times \mathcal{Y}$ and $z \in \mathcal{Z}$, $F(z|x^c, \eta)$ is Lipschitz continuous in $x^c \in \mathcal{X}^c$ and has all partial derivatives up to order 2 with respect to x^c.

(ii) For each $\eta \in \mathcal{X}^d \times \mathcal{Y}$ and $z \in \mathcal{Z}$, the second order partial derivatives with respect to x^c, $\partial^2 F(z|x^c, \eta)/\partial x_s^c \partial x_t^c$, $s, t = 1, \cdots, p_c$, are uniformly bounded and Hölder continuous on $\mathcal{X}^c$: for x^c, $\tilde{x}^c \in \mathcal{X}^c$, $|\partial^2 F(z|x^c, \eta)/\partial x_s^c \partial x_t^c - \partial^2 F(z|\widetilde{x}^c, \eta)/\partial x_s^c \partial x_t^c| \leq C \; ||x^c - \widetilde{x}^c||$, where C is a generic finite constant and $\|\cdot\|$ denotes the Euclidean norm.

(iii) For each $x^c \in \mathcal{X}^c$ and $\eta \in \mathcal{X}^d \times \mathcal{Y}$, $|F(z|x^c, \eta) - F(\widetilde{z}|x^c, \eta)| \leq C\,|z - \widetilde{z}|$ for all $z, \widetilde{z} \in \mathcal{Z}$.

Assumption A.4 (i) The kernel function $q : \mathbb{R} \to \mathbb{R}^+$ is a continuous, bounded, and symmetric PDF.

(ii) $u \to |u|^4 q(u)$ is integrable on $\mathbb{R}$ with respect to Lebesgue measure.

(*iii*) Let $\mathbf{q}_j(u) \equiv u^j q(u)$ for all $j = 0, \cdots, 3$. For some $C_1 < \infty$ and $C_2 < \infty$, either $q(\cdot)$ is compactly supported such that $q(u) = 0$ for $|u| > C_1$, and $|\mathbf{q}_j(u) - \mathbf{q}_j(\widetilde{u})| \leq C_2\,|u - \widetilde{u}|$ for any $u, \widetilde{u} \in \mathbb{R}$ and for all $j = 0, \cdots, 3$; or $q(\cdot)$ is differentiable, $|d\mathbf{q}_j(u)/du| \leq C_1$, and for some $\iota_0 > 1$, $|d\mathbf{q}_j(u)/du| \leq C_1 |u|^{-\iota_0}$ for all $|u| > C_2$ and for all $j = 0, \cdots, 3$.

Assumption A.5 Let $h! \equiv \Pi_{s=1}^{p_c} h_s$. As $n \to \infty$, $\|h\| \to 0$, $\|\lambda\| \to 0$, $||\lambda||$ *is of the same order as* $||h||^2$, $n\,(h!)^2/\log n \to \infty$, $n\,(h!)^{1/2}\|h\|^4 \to 0$, and $\|h\|^4/h! \to 0$.

Remark 1. The IID assumption in assumption A.1 is standard in cross sectional study. One could allow heterogeneity but that would complicate the presentation to a large degree. Assumption A.2 is standard for nonparametric local polynomial estimation with mixed regressors. Assumptions A.3-A.4 are used to obtain uniform consistency for the local polynomial estimator of Masry (1996) and Hansen (2008). Assumption A.5 imposes appropriate conditions on the bandwidth. In particular A.5 implies that undersmoothing is required for our test and $p_c < 4$. This is typical in nonparametric tests when local linear regression is involved. In the case where $p_c \geq 4$, one has to rely upon higher order local polynomial regressions.

3.4.3 The Asymptotic Distribution of the Test Statistic

Let $\overline{\mathbf{S}}_y(x) \equiv E[K_{h\lambda}(X_j, x)\mathbf{1}_j^y \boldsymbol{\tau}_h(X_j^c - x^c)\boldsymbol{\tau}_h(X_j^c - x^c)'], \mathbf{K}_y(\zeta_j, x) \equiv e_1'[\overline{\mathbf{S}}_y(x)]^{-1}\boldsymbol{\tau}_h\left(X_j^c - x^c\right)$ $K_{h\lambda}(X_j, x)\mathbf{1}_j^y$, and $\overline{\mathbf{1}}_{z,y}(\varsigma_i) \equiv \mathbf{1}\{Z_i \leq z\} - F(z|X_i, y)$. Define

$$B_n \equiv \frac{(h!)^{1/2}}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1}\left[\mathbf{K}_r\left(\zeta_j, X_i\right)\overline{\mathbf{1}}_{Z_i,r}(\varsigma_j) - \mathbf{K}_s\left(\zeta_j, X_i\right)\overline{\mathbf{1}}_{Z_i,s}(\varsigma_j)\right]^2, \tag{4.9}$$

and

$$\sigma_n^2 \equiv 2h!E_iE_j\left[\sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1}\int\left\{\mathbf{K}_r(\zeta_i;x)\overline{\mathbf{1}}_{z,r}(\varsigma_i) - \mathbf{K}_s(\zeta_i,x)\overline{\mathbf{1}}_{z,s}(\varsigma_i)\right\}\{\mathbf{K}_r\left(\zeta_j;x\right)\overline{\mathbf{1}}_{z,r}(\varsigma_j) - \mathbf{K}_s(\zeta_i,x)\overline{\mathbf{1}}_{z,s}(\varsigma_i)\}F_\xi(d\xi)\right]^2 \tag{4.10}$$

where E_i denote the expectation with respect to ξ_i. Let $\sigma_0^2 \equiv \lim_{n\to\infty}\sigma_n^2$.

Our first result says that after centering, $(h!)^{1/2} D_n$ is asymptotically normally distributed under H_0.

Theorem 3.4.1 *Suppose Assumptions A.1-A.5 hold. Then under* H_0, $(h!)^{1/2} D_n - B_n \xrightarrow{d} N\left(0, \sigma_0^2\right)$.

To implement the test, we need to consistently estimate B_n and σ_0^2. For this purpose, let $\widehat{\mathbf{1}}_{Z_i,y}(\varsigma_j) \equiv \mathbf{1}\{Z_j \leq Z_i\} - \widehat{F}(Z_i|X_j, y)$.
Let $\widehat{\mathbf{K}}_r\left(\zeta_j; x\right) \equiv e_1'[\mathbf{S}_{nr}(x)]^{-1}\boldsymbol{\tau}_h\left(X_j^c - x^c\right)K_{h\lambda}(X_j, x)\mathbf{1}_j^y$. Let

$$\widehat{\alpha}_{ij,rs} \equiv \widehat{\mathbf{K}}_r\left(\zeta_j; X_i\right)\widehat{\mathbf{1}}_{Z_i,r}(\varsigma_j) - \widehat{\mathbf{K}}_s\left(\zeta_j; X_i\right)\widehat{\mathbf{1}}_{Z_i,s}(\varsigma_j), \text{ and } \widehat{\beta}_{ij,rs} \equiv \frac{1}{n}\sum_{l=1}^{n}\widehat{\alpha}_{li,rs}\widehat{\alpha}_{lj,rs}.$$

Define

$$\widehat{B}_n \equiv \frac{(h!)^{1/2}}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1}\widehat{\alpha}_{ij,rs}^2, \text{ and } \widehat{\sigma}_n^2 \equiv \frac{2h!}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left[\sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1}\widehat{\beta}_{ij,rs}\right]^2.$$

We demonstrate in Theorem 3.4.2 below that $\widehat{B}_n - B_n = o_P(1)$ and $\widehat{\sigma}_n^2 - \sigma_0^2 = o_P(1)$. Then we can compare

$$T_n \equiv \left((h!)^{1/2} D_n - \widehat{B}_n\right)/\sqrt{\widehat{\sigma}_n^2} \tag{4.11}$$

to the one-sided critical value z_α, the upper α percentile from the $N(0,1)$ distribution. We reject the null at level α if $T_n > z_\alpha$.

To examine the asymptotic local power of the test, we consider the following sequence of Pitman local alternatives:

$$H_1(\gamma_n) : F(z|x,r) - F(z|x,s) = \gamma_n \delta_{n,rs}(\varsigma) \text{ for a.e. } \xi, \tag{4.12}$$

where $\gamma_n \to 0$ as $n \to \infty$ and $\delta_{n,rs}(\cdot)$ is a continuous function such that $\mu_0 \equiv \lim_{n\to\infty} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} E[\delta_{n,rs}(\varsigma_i)]^2 < \infty$. The following theorem establishes the local power of the test.

Theorem 3.4.2 *Suppose Assumptions A.1-A.5 hold.*
Then under $H_1(\gamma_n)$ with $\gamma_n = n^{-1/2}(h!)^{-1/4}$, $T_n \xrightarrow{d} N(\mu_0/\sigma_0, 1)$.

Thus, the test has nontrivial power against Pitman local alternatives that converge to zero at rate $n^{-1/2}(h!)^{-1/4}$. The asymptotic local power function is given by $1 - \Phi(z_\alpha - \mu_0/\sigma_0)$, where Φ is the standard normal CDF.

The following theorem establishes the consistency of the test under the global alternative H_1 stated in (4.2).

Theorem 3.4.3 *Suppose Assumptions A.1-A.5 hold. Then under H_1, $n^{-1}(h!)^{-1/2}T_n = \mu_A/\sigma_0 + o_P(1)$, where $\mu_A \equiv \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} E[F(Z_i|X_i,r) - F(Z_i|X_i,s)]^2$, so that $P(T_n > c_n) \to 1$ under H_1 for any nonstochastic sequence $c_n = o\left(n(h!)^{1/2}\right)$.*

Remark 2. Alternatively, one can consider testing the conditional independence of Y and Z given X based upon the comparison of $F(z|x)$ with $F(z|x,y)$. In this case, the test statistic would be

$$\widetilde{D}_n \equiv \sum_{i=1}^{n} \left[\widetilde{F}(Z_i|X_i) - \widetilde{F}(Z_i|X_i,Y_i)\right]^2,$$

where $\widetilde{F}(z|x)$ and $\widetilde{F}(z|x,y)$ are local linear estimates of $F(z|x)$ and $F(z|x,y)$ by smoothing all discrete variables in X_i and (X_i, Y_i), respectively. After being suitably centered and rescaled, $\widetilde{D}_n$ can be shown to be asymptotically normally distributed. The key assumption for the asymptotic normality of $\widetilde{D}_n$ would require that the bandwidth (λ_y, say) used in smoothing the discrete variable Y_i tends to zero as $n \to \infty$. Nevertheless, under the null hypothesis of conditional independence, Y_i is an irrelevant variable in the prediction of Z_i or $\mathbf{1}(Z_i \le z)$, implying that the optimal bandwidth for λ_y should tend to 1 as $n \to \infty$ (see Li and Racine (2007)). Thus this creates a dilemma for the choice of λ_y, making it extremely difficult to control the finite sample level of a test based upon $\widetilde{D}_n$. In contrast, when we construct our D_n test statistic, we obtain the estimate $\widehat{F}(Z_i|X_i,y)$ of $F(Z_i|X_i,y)$ for different values of y without smoothing the discrete variable Y_i (see (4.8)) and thus avoid the above dilemma.

3.5 Monte Carlo Simulations

In this section we conduct some Monte Carlo experiments to evaluate the finite sample performance of our test. We consider two data generating processes (DGPs):

DGP 1.

$$\begin{aligned} Y_i &= \mathbf{1}\left(\varepsilon_{Yi} \le m_Y\left(X_i\right)\right), \\ Z_i &= \mathbf{1}\left(\varepsilon_{Zi} \le m_Z\left(X_i\right)\right), \\ m_Y\left(X_i\right) &= \frac{X_{i1}^c - 0.5X_{i2}^c + \phi\left(X_{i2}^c\right) - X_{i1}^c X_{i2}^c - 0.5X_{i1}^c X_{i1}^d + 0.5X_{i1}^d + 0.5X_{i1}^d X_{i2}^d}{\sqrt{1 + X_{i1}^{c2} + X_{i2}^{c2}}}, \\ m_Z\left(X_i\right) &= \phi\left(X_{i1}^c\right) X_{i2}^c - X_{i1}^c - X_{i2}^c X_{i2}^d + 0.5X_{i1}^d X_{i2}^d + \delta Y_i X_{i1}^c, \end{aligned}$$

where $X_i \equiv \left(X_{i1}^c, X_{i2}^c, X_{i1}^d, X_{i2}^d\right)'$, ϕ is the $N(0,1)$ PDF, X_{i1}^c is IID $U(0,4)$, X_{i2}^c is IID, computed as the sum of 48 independent random variables, each uniformly distributed on $[-0.25,\ 0.25]$, $P\left(X_{i1}^d = l\right) = 1/4$ for $l = 0, 1, 2, 3$, $P\left(X_{i2}^d = l\right) = 1/5$ for $l = 0, 1, 2, 3, 4$, ε_{Y1} is IID $N(0,1)$, ε_{Zi} is IID $N(0,1)$, and all these variables are mutually independent. δ controls the degree of conditional dependence between Y_i and Z_i given X_i. Given X_i, Y_i and Z_i are conditionally independent when $\delta = 0$ and conditionally dependent otherwise.

DGP 2.

$$\begin{aligned} Y_i &= \mathbf{1}\left(\varepsilon_{Yi} \le m_Y\left(X_i\right)\right), \\ Z_i &= m_Z\left(X_i\right) + s\ \varepsilon_{Zi}, \end{aligned}$$

where $X_i \equiv \left(X_{i1}^c, X_{i2}^c, X_{i1}^d, X_{i2}^d\right)'$, ε_{Yi} and ε_{Zi} are generated as in DGP 1, m_Y and m_Z are as defined in DGP1, and s is taken to ensure the signal-noise ratio in the equation for Z_i to be 1 across all simulations.

Clearly, DGP 1 generates binary Y_i and Z_i variables whereas DGP 2 generates binary Y_i and continuous Z_i. In both DGPs the X_i vector includes two continuous variables, X_{i1}^c and X_{i2}^c, and two discrete variables, X_{i1}^d and X_{i2}^d. Note that our test is based on local linear regressions, which typically require compactly supported conditioning variables. This motivates the otherwise awkward way we generate X_{i2}^c in DGPs 1-2. According to the central limit theorem, we can treat X_{i2}^c as being nearly standard normal random variables but with compact support $[-12,\ 12]$.

Notice that the two discrete variables in X_i partition the data into $4 \times 5 = 20$ cells. In conjunction with the 2 categories of dummy Y_i, this will partition the data into $20 \times 2 = 40$ cells if we adopt the conventional *nonparametric frequency approach* to do the estimation and testing. If the number of observations n is small, say, 100, each cell has a tiny amount of observations on average and some empty cells in practice, this will make the estimation of the CDF $F(z|x,y)$ extremely difficult. A nonparametric-frequency-based test should not be expected to perform well in terms of both level and power. Even with nonparametric smoothing over the discrete variables in X_i as advocated by our test, the problem continues to be hard but less severe.

To construct the test statistic, we need to choose both kernel and bandwidth. We choose the product of Gaussian kernel for the two continuous regressors: $q(x) = (2\pi)^{-1/2} \exp\left(-x^2/2\right)$. Since there is no data-driven procedure to choose the bandwidths $h = (h_1, h_2)'$ and $\lambda = (\lambda_1, \lambda_2)'$ for our testing problem, we choose them according to the rule of thumb:

$$h_l = \gamma s_{X_l^c} n^{-1/4.5},\ \lambda_l = \gamma n^{-2/4.5} \text{for } l = 1, 2, \tag{5.1}$$

where $s_{X_l^c}$ is the sample standard deviation of X_{il}^c and γ is a fixed constant. We study the behavior of our test with different choices of γ in order to examine the sensitivity of our test to the bandwidth sequence. Robinson (1991, p.448) proposes very similar devices. Note that these choices for h and λ and the kernel function meet the requirements for our test. Through a preliminary simulation study, we find our bootstrap-based test is not sensitive to the choice of γ when we take $\gamma \in [0.5, 2]$. So we fix $\gamma = 1$ for our simulation results.

It is well known that the asymptotic normal distribution typically cannot approximate the finite sample distribution of many nonparametric test statistics. This is especially true for our test when we have discrete conditioning variables in X_i with reasonably large number of categories. So we suggest using a bootstrap method to obtain the bootstrap p-values. Here, we generate the bootstrap data $\{(X_i^*, Y_i^*, Z_i^*)\}_{i=1}^n$ based on the following local bootstrap procedure:

1. Set $(X_i^*, Y_i^*) = (X_i, Y_i)$ for each $i \in \{1, \cdots, n\}$.

2. For $i = 1, \cdots n$, given X_i^*, draw Z_i^* from the following local constant nonparametric estimate of $F(z|X_i^*)$:

$$\widetilde{F}_{\widetilde{h}\widetilde{\lambda}}(z|X_i^*) = \frac{\sum_{j=1}^n K_{\widetilde{h}\widetilde{\lambda}}(X_j, X_i^*) \mathbf{1}(Z_j \leq z)}{\sum_{j=1}^n K_{\widetilde{h}\widetilde{\lambda}}(X_j, X_i^*)} \tag{5.2}$$

   where $\widetilde{h}$ and $\widetilde{\lambda}$ are the bandwidth used in the estimation of $F(z|X_i^*)$.

3. Compute the bootstrap test statistic T_n^* in the same way as T_n by using $\{(X_i^*, Y_i^*, Z_i^*)\}_{i=1}^n$ instead.

Table 3.1: Finite sample rejection frequency for DGPs 1-2

DGP	Sample size n	δ	Our test $h_l = s_{X_l^c} n^{-1/4.5},\ \lambda_l = n^{-2/4.5}$ 1%	5%	10%	Nonparametric frequency approach $h_l = s_{X_l^c} n^{-1/4.5},\ \lambda_l = 0$ 1%	5%	10%
1	200	0	0.036	0.092	0.144	0.100	0.224	0.288
		1	0.420	0.592	0.644	0.200	0.268	0.348
		2	0.900	0.920	0.924	0.472	0.620	0.656
	400	0	0.024	0.068	0.104	0.088	0.216	0.296
		1	0.628	0.720	0.764	0.156	0.312	0.384
		2	0.904	0.924	0.932	0.512	0.648	0.728
	800	0	0.016	0.040	0.068	0.160	0.348	0.448
		1	0.840	0.880	0.884	0.204	0.396	0.492
		2	0.968	0.968	0.968	0.628	0.752	0.836
2	200	0	0.020	0.068	0.124	0.028	0.080	0.176
		1	0.064	0.160	0.268	0.068	0.172	0.232
		2	0.176	0.300	0.448	0.088	0.240	0.312
	400	0	0.004	0.016	0.072	0.032	0.132	0.196
		1	0.080	0.192	0.288	0.068	0.172	0.248
		2	0.268	0.504	0.632	0.100	0.196	0.328
	800	0	0.000	0.032	0.056	0.020	0.128	0.216
		1	0.168	0.304	0.408	0.100	0.276	0.360
		2	0.600	0.768	0.812	0.136	0.296	0.380

4. Repeat steps 1-3 B times to obtain B bootstrap test statistic $\left\{T_{nj}^*\right\}_{j=1}^B$. Calculate the bootstrap p-values $p^* \equiv B^{-1}\sum_{j=1}^B \mathbf{1}\left(T_{nj}^* \geq T_n\right)$ and reject the null hypothesis of conditional independence if $p^*$is smaller than the prescribed level of significance.

The above procedure is coined as the local bootstrap procedure by Paparoditis and Politis (2000) who also explain how to generate the bootstrap observations computationally. It works no matter whether Z_i is discrete or continuous. In the case where Z_i is continuous, we can also generate a smooth version of Z_i^* through $Z_i^{**} = Z_i^* + b\eta_i$, where $b \equiv b(n) \to 0$ as $n \to \infty$, and η_i is drawn from $N(0,1)$. In our simulations and applications, we generate Z_i^* and Z_i^{**} for the case where Z_i is discrete and continuous, respectively. When Z_i is continuous, we set $b = s_Z n^{-1/6}$ with s_Z being the sample standard deviation of Z_i. Our simulations indicate that the choice of b plays little role in the performance of our test. For simplicity, we set $\widetilde{h} = h$ and $\widetilde{\lambda} = \lambda$.

Table 3.1 reports the empirical rejection frequencies of our test at 1%, 5%, and 10% nominal levels for DGPs 1-2. Also reported in the table is a variant of our test based on the idea of nonparametric frequency, which is obtained by setting the smoothing parameters for the discrete variables in X_i to be 0 in the calculation of our test statistic. To save on computational time, we use 250 replications for each sample size n and 100 bootstrap resamples in each replication. We summarize some important findings from Table 3.1.

First, the level of our nonparametric smoothing test is reasonably well behaved despite the fact that it tends to be oversized when n is small and the average number of observations per cell is small. In the case where $n = 200$, $X_i^d \equiv (X_{i1}^d,\ X_{i2}^d)'$ and Y_i partition the 200 observations into 40 cells so that each cell contains only 5 observations on average. Given the two conditioning variables X_{i1}^c and X_{i2}^c, one cannot expect the conditional CDF for each cell values of X_i^d and Y_i to be well estimated no matter whether we choose to smooth X_i^d or not. This definitely has some adverse effect on the performance of our test. Despite this, our nonparametric smoothing test seems to perform well even if n is small and the average number of observations per cell is small. As n and the average number of observations per cell double, the levels of our test tend to be improved and get close to the nominal levels.

Second, our test has power to detect deviations from conditional independence no matter whether Z_i is discrete or continuous. In DGP 1 when δ changes from 0 to 1 (resp. 2) so that Y_i becomes to affect Z_i conditional on X_i, the unconditional probability for Z_i to take value 1 increases from 0.38 to 0.52 (resp. 0.60). Our nonparametric smoothing test can detect such changes very well even for small n. As n doubles, the above changes of unconditional probabilities remain the same as we change δ, but the power of our test increases. In DGP 2, Z_i is continuously valued. The power performance does not appear to be as well as the case of DGP 1 because we normalize the error terms in the equation for Z_i to ensure the signal-noise ratio to be 1 across different values of δ. If we set $s = 1$ in the equation for Z_i and allow the signal to become stronger as δ increases, we can observe significant improvement of the power performance of our test.

Third, in terms of both size and power, our smoothing nonparametric test significantly dominates the nonparametric-frequency-based test. The latter test tends to be oversized for both DGPs and all sample sizes under investigation. Despite its oversize, as expected, the latter test is much less powerful in detecting deviations from the null of conditional independence than our nonparametric smoothing test.

3.6 Empirical Application

In this section we apply the nonparametric test to an automobile insurance data set.[9] We first briefly introduce the automobile insurance market in France where our data set stems from, then discuss configurations of the data set and present our empirical findings. Noting that the design of automobile insurance is relatively similar in most countries, so we believe that our methodology is broadly applicable.

3.6.1 Principles of the Automobile Insurance in France

In France, like in many other countries, all cars must be insured at the "responsabilité civile" (RC) level.[10] This is a liability insurance that covers damage inflicted to other drivers and their cars. Moreover, insurance companies offer additional non-compulsory coverage. The most common one is called "assurance tous risques" (TR), which also covers damage to the insured car or the driver in the case of an accident at which he or she is at fault. The insurees can choose from different comprehensive insurance contracts which vary in the value of the deductible (fixed or proportional).

A special feature of the car insurance is the so called "bonus/ malus", a uniform experience rating system. At any date/year t, the premium is defined as the product of a basis amount and a "bonus" coefficient. The basic amount can be defined freely by the insurance companies according to their risk classification but cannot be related to past experience. The past experience is captured by the so called "bonus/ malus" coefficient whose evolution is strictly regulated. Suppose, the bonus coefficient is b_t at the beginning of the tth period. Then the occurrence of an accident during the period leads to an increase of 25 percent at the end of the period (i.e., $b_{t+1} = 1.25b_t$), whereas an accident-free year implies a reduction of 5 percent at the end (i.e., $b_{t+1} = 0.95b_t$). Additionally, several special rules are applied, which include the permission to overcharge contracts held by young drivers. But the surcharge is limited to 140 percent of the basis rate and is forced to decrease by half every year in which the insuree has not had an accident.

The basis amount of the premium is calculated according to different risk classes. Due to variables like age, sex, profession, area, etc., the insurees are divided into different risk classes which should reflect their accident probabilities, and the premium to be paid is then determined.

[9]Despite the scarcity of insurance data sets the car insurance has been analyzed for different countries amongst others by Chiappori and Salanié (1997, 2000), Richaudeau (1999), Cohen (2005), Saito (2006) and Kim et. al. (2009).

[10]The description of the French car insurance follows Chiappori and Salanié (2000).

3.6.2 Configurations of the Data Set

We use a data set of the French federation of insurers (FFSA) which conducted in 1990 a survey of its members.[11] This data set was also used in Chiappori and Salanié (1997, 2000). With a sampling rate of 1/20 the data set consists of 41 variables on $1,120,000$ contracts and 25 variables on $120,000$ accidents for the year 1989. For each driver all variables which are used by insurance companies for pricing their contracts - age of the driver, sex, profession of the driver, year of drivers license, age of the car, type of the car, use of the car, and area - plus the characteristics of the contract and the characteristics of the accident, if occurred, are available. We restrict our analysis to all "young"drivers who obtained their driver license in 1988.[12] This reduces the sample size to $n = 6,333$.

As Chiappori and Salanié (2000) argue, focusing on young drivers has two major advantages. In a subsample of young drivers the driving experience is much more homogeneous than that in the total population in which groups of different experiences are pooled. Therefore the heteroskedasticity problem is mitigated and less severe. The concentration on young drivers also avoids the problems associated with the experience rating and the resulting bias. The past driving history is usually observed by the insurance companies. The past driving records are highly informative on probabilities of accident and used for pricing. The bonus coefficient is a very excellent proxy for this variable. However, the introduction of this variable is quite delicate because of its endogeneity. This problem can be circumvented either by using panel data or by using only data on beginners.[13] We pursue the second approach and concentrate on novice drivers.

One important issue in testing for asymmetric information is the distinction between accidents and claims.[14] The data set of insurance companies comprises claims. But whether an accident - once it has occurred - is declared to the insurance company and becomes a claim depends on the decision of the insuree. This decision is mainly determined by the nature of the contract. For example, accidents whose damage is below the deductible or is not covered are usually not declared. Therefore one might expect a positive correlation between the type of contract (coverage) and the probability of a claim - even in the absence of ex ante moral hazard.[15] One strategy to handle this problem is to discard all accidents in which only one automobile was involved. Whenever two cars are involved, a declaration is nearly inevitable.

[11] The FFSA comprehends 21 companies that together have 70 percent market share of the French automobile insurance market.

[12] "Young" refers not to the actual age but to the driving experience.

[13] For a detailed discussion see Chiappori and Heckmann (1999).

[14] This problem is, e.g., discussed in detail in Cohen and Siegelman (2010).

[15] The phenomenon that accidents that are not covered are not declared is sometimes called "ex post moral hazard".

To make the results comparable with those of Chiappori and Salanié (2000) and to check for robustness we examine several different configurations of the data set. Let X_i denote the set of exogenous control variables for individual i. Let $Y_i = 0$ if individual i buys only the minimum legal coverage (a RC contract) and 1 if individual i buys any form of comprehensive coverage (a TR contract). First we consider discrete Z_i where $Z_i = 1$ if i has at least one accident in which he or she is judged to be at fault and 0 otherwise (no accident occurred or i was not at fault). Then we consider the case where Z_i is continuous and defined by the total payments caused by the insuree, which is also included in the data set.

For the random variables in X_i, we consider three configurations. In Configuration I we include the following control variables in X_i : sex (2), make of car (8), performance of the car (6), type of use (4), type of area (5), profession of the driver (8), region (10), age of the driver, and age of the car, where numbers in brackets indicate the number of categories for the corresponding discrete variables, and variables without numbers indicate they are continuous variables. These control variables are similar to those used by Chiappori and Salanié (2000) for their probit-model- or χ^2-based tests except that we do not transform the age of the car and that of the driver to discrete variables.

Our nonparametric test requires that the number of observations per cell should not be too small. So we also consider another two configurations for X_i. In Configuration II we omit the variable, make of the car, which describes the home country of the manufacturer of the car. We think that the most important part of the information concerning an automobile can be captured by the performance of the car, so that the omission of this variable should have no significant influence on the results. For example, the accident probability of an Italian and a French compact car should not differ significantly, all other things being equal. Additionally, we reduce the number of categories for some discrete variables according to Column 3 in Table 3.2. Again, we argue that merging categories which are nearly identical or closely related does not bias the results.

In Configuration III we use only two categories for each of the seven discrete variables in X_i. As surveyed above, Salanié and Chiappori (2000) also conduct nonparametric tests where they code all control variables as binary and apply a χ^2 independence test to each cell, and then aggregate the resulting test statistics in three different ways. Our third configuration enables a direct comparison of our nonparametric test with their nonparametric tests.

Configurations IV - VI correspond to Configurations I - III, respectively. In the settings IV - VI we only replace the discrete dummy variable Z_i by its continuous counterpart, i.e., by the total payments caused through accidents by the insuree to the insurance company. In all configurations, we treat the age of the car and the age of the driver as continuous variables.

See Table 3.2 for a summary of these configurations.

Table 3.2: An overview of the data configurations

Variables\Configurations	I	II	III	IV	V	VI
Y_i	2	2	2	2	2	2
Z_i	2	2	2	X	X	X
sex	2	2	2	2	2	2
make of car	8	-	2	8	-	2
performance of car	6	6	2	6	6	2
type of use	4	3	2	4	3	2
type of area	5	2	2	5	2	2
profession of driver	9	5	2	9	5	2
region	10	5	2	10	5	2
age of driver	X	X	X	X	X	X
age of car	X	X	X	X	X	X

Note: Integers denote the number of categories for the corresponding discrete variables. An "X" in the table denotes that the corresponding variable is a continuous variable.

3.6.3 Empirical Results

In this subsection we apply the nonparametric test to the data set introduced in the above subsection. Table 3.3 reports the bootstrap p-values for our nonparametric test under various configurations of the data set. Given the large sample size ($n = 6,333$) and the need of bootstrap, the computational burden for our bootstrap-based nonparametric test is very heavy. Simulations for smaller sample sizes with different choices of the number of bootstrap replications ($B = 100$, 200, 300) indicate that our testing results are insensitive to the choice of B. So we only set $B = 100$ for our applications. Also due to the large sample size and the large number of control variables, it is difficult to use least squares cross validation method to choose data-driven bandwidths to conduct our nonparametric test. For this reason we adopt the rule of thumb to choose the bandwidths: $h_l = \gamma s_{X_l^c} n^{-1/4.5}$ and $\lambda_s = \gamma n^{-2/4.5}$ for the continuous and discrete control variables, respectively. To check the sensitivity of the test to the choice of bandwidth, we consider four values of γ : 0.75, 1, 1.25, and 1.5. These choices of bandwidths fulfill the requirements of Assumption A.5.

We summarize some important findings from Table 3.3. First, it indicates that in all cases we fail to reject the null hypothesis of absence of asymmetric information at the 10% significance level. This means that the knowledge of the choice of the contract does not contain

Table 3.3: Bootstrap p-values for our nonparametric test under various configurations

γ\Configurations	I	II	III	IV	V	VI
$\gamma = 0.75$	0.82	0.54	0.24	1.00	1.00	1.00
$\gamma = 1$	0.79	0.62	0.13	1.00	1.00	1.00
$\gamma = 1.25$	0.77	0.66	0.14	1.00	1.00	1.00
$\gamma = 1.5$	0.76	0.72	0.21	1.00	1.00	1.00

information for predicting the probability of an accident or the other way round, knowing the number of accidents (discrete) or the caused damages (continuous) is of no value for predicting the chosen contract. Therefore our test affirms the Chiappori and Salanié's (2000) findings that there is no evidence of asymmetric information in the market for automobile insurance in France. The results are very robust to different configurations of data and choices of bandwidth. Second, Table 3.3 reveals that an aggregation of the categories of the discrete control variables leads to a decrease of the p-values so that a reduction of information might disguise asymmetric information. Therefore, a (non-)parametric test that relies on highly aggregated information might yield misleading or wrong conclusions. Third, Table 3.3 also reveals that using the payments of the insurance companies instead of the number of accidents leads to a strengthening of the absence of asymmetric information. Again, a reduction of information might lead to wrong test conclusions.

Recently Kim et al. (2009) have argued that the absence of asymmetric information in most empirical studies might be due to the dichotomous measurement approach that induces the excessive bundling of contracts with different deductibles. In reality the insurees can choose between several deductibles referring to different fields of coverage. But most studies aggregate this choice opportunities to a binary choice between "compulsory" coverage and "additional" coverage so that the choice variable Y_i becomes binary. Kim et al. (2009) claim that excessive bundling in coverage measurements might disguise the existence of asymmetric information. So they apply a multinomial measurement approach, which is parametric in nature, and demonstrate the evidence of asymmetric information in their data set obtained from a major automobile insurance company in Korea.

Since our data set also contains the exact level of the chosen deductible, we can investigate this hypothesis as our test is fully applicable to this problem. A very small proportion of the contracts has proportional deductibles which are dropped for this analysis. Therefore the sample size decreases to $n = 6,219$. We divide the chosen deductible into three ($0 - 100$, $101 - 1500$, and > 1500) groups. The results are reported in Table 3.4 for different data

Table 3.4: Bootstrap p-values for our nonparametric test when the choice variable has three deductible levels

γ\Configurations	I	II	III
$\gamma = 0.75$	0.71	0.73	0.73
$\gamma = 1$	0.67	0.71	0.88
$\gamma = 1.25$	0.69	0.64	0.68
$\gamma = 1.5$	0.64	0.77	0.75

configurations introduced above. In comparison to the settings defined in Table 3.2, the choice variable Y_i now has three categories, but everything else remains unchanged in the data set. Clearly, Table 3.4 confirms the absence of asymmetric information in the data. We also tried a finer division for the deductible so that Y_i has more categories. In all cases, our results are robust in that they all confirm the absence of asymmetric information in the data. Intuitively speaking, if there is no asymmetric information in the most important choice between compulsory and comprehensive insurance, one should not expect asymmetric information in the minor decision of the exact deductible when the money at stack is not so high.

3.7 Concluding Remarks

We propose a new nonparametric test for asymmetric information in this paper and apply it to a French automobile insurance data set. Our main conclusion is that we cannot detect asymmetric information in the data set despite different configurations of the control variables and different choices of bandwidth parameters. Our nonparametric test does not require specification of any functional or distributional form among the sets of variables of interest and it is not subject to any misspecification problem given the right choice of control variables. We also show that excessive bundling does not necessarily result in a disguise of asymmetric information. Both in the case of the binary choice between "compulsory" coverage and "additional" coverage and in the case of several deductibles (three and more groups) we confirm the absence of asymmetric information. Our results are also very strong in contrast to Kim et al. (2009).

Since nearly all other classes of insurance, such as the legal protection insurance, private health insurance, and disability insurance, are structured in the same way as the auto insurance, applications to data sets in these subfields are immediate and might help to gain new insights. Moreover, our test can be applied to more general settings, either to testing for asymmetric information in other fields or more generally, to testing the general hypothesis of conditional independence.

3.8 Mathematical Appendix

Let $\Delta_{j,y}(x,z) \equiv F(z|X_j,y) - F(z|x,y) - \sum_{s=1}^{p_c} \left(\partial F\left(z|x^c,x^d,y\right)/\partial x_s^c\right)\left(X_{js}^c - x_s^c\right)$, $\mathbf{V}_{ny}(\varsigma) \equiv \frac{1}{n}\sum_{j=1}^{n} K_{h\lambda}(X_j,x)\,\mathbf{1}_j^y \boldsymbol{\tau}_h(X_j^c - x^c)\overline{\mathbf{1}}_{z,y}(\varsigma_j)$, and
$\mathbf{B}_{ny}(\varsigma) \equiv \frac{1}{n}\sum_{j=1}^{n} K_{h\lambda}(X_j,x)\,\mathbf{1}_j^y \boldsymbol{\tau}_h(X_j^c - x^c)\Delta_{j,y}(x,z)$. Let $\overline{\mathbf{S}}_y(x) \equiv E[\mathbf{S}_{ny}(x)]$ and $\overline{\mathbf{B}}_y(\varsigma) \equiv E[\mathbf{B}_{ny}(\varsigma)]$. The following lemma establishes the uniform consistency of $\widehat{F}(z|x,y)$.

Lemma 3.8.1 *Suppose Assumptions A.1-A.5 hold. Then uniformly in* $\xi \equiv (x',y,z)'$ *we have:* $\widehat{F}(z|x,y) - F(z|x,y) = e_1'[\overline{\mathbf{S}}_y(x)]^{-1}[\mathbf{V}_{ny}(\varsigma) + \overline{\mathbf{B}}_y(\varsigma)] + O_P(\nu_n^2 + \nu_n(\|h\|^2 + \|\lambda\|)) = O_P(\nu_n + \|h\|^2 + \|\lambda\|)$, *where* $\nu_n \equiv n^{-1/2}(h!)^{-1/2}\sqrt{\log n}$.

Proof. Since $[\mathbf{S}_{ny}(x)]^{-1}\ \mathbf{S}_{ny}(x) = I_{p_c+1}$ where I_{p_c+1} is a $(p_c+1)\times(p_c+1)$ identity matrix, we obtain the following standard bias and variance decomposition:

$$\widehat{F}(z|x,y) - F(z|x,y) = e_1'[\mathbf{S}_{ny}(x)]^{-1}\mathbf{V}_{ny}(\varsigma) + e_1'[\mathbf{S}_{ny}(x)]^{-1}\mathbf{B}_{ny}(\varsigma), \tag{8.1}$$

where e_1' is the first row of I_{p_c+1}. By Theorems 2 and 4 in Masry (1996) with little modification to account for discrete regressors,[16]

$$\mathbf{S}_{ny}(x) = \overline{\mathbf{S}}_y(x) + O_P(\nu_n),\ \mathbf{V}_{ny}(\varsigma) = O_P(\nu_n),\ \text{and}\ \mathbf{B}_{ny}(\varsigma) - \overline{\mathbf{B}}_y(\varsigma) = O_P(\nu_n(\|h\|^2 + \|\lambda\|)),$$

where the probability orders hold uniformly in $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. By the Slutsky lemma,

$$[\mathbf{S}_{ny}(x)]^{-1} = \left\{\overline{\mathbf{S}}_y(x) + \left[\mathbf{S}_{ny}(x) - \overline{\mathbf{S}}_y(x)\right]\right\}^{-1} = [\overline{\mathbf{S}}_y(x)]^{-1} + O_P(\nu_n). \tag{8.2}$$

By the same argument as used in the proof of Theorem 4.1 of Boente and Fraiman (1991), we can show that $\mathbf{V}_{ny}(\varsigma) = O_P(\nu_n)$ uniformly in ς under Assumption A.3. It follows that $\widehat{F}(z|x,y) - F(z|x,y) = e_1'\{[\overline{\mathbf{S}}_y(x)]^{-1} + O_P(\nu_n)\}\{\mathbf{V}_{ny}(\varsigma) + [\overline{\mathbf{B}}_y(\varsigma) + O_P(\nu_n(\|h\|^2 + \|\lambda\|)]\}$ $= e_1'[\overline{\mathbf{S}}_y(x)]^{-1}[\mathbf{V}_{ny}(\varsigma) + \overline{\mathbf{B}}_y(\varsigma)] + O_P(\nu_n^2 + \nu_n(\|h\|^2 + \|\lambda\|)) = O_P(\nu_n + \|h\|^2 + \|\lambda\|)$. ■

Proof of Theorems 3.4.1 and 3.4.2

We only prove Theorem 3.4.2, as the proof of Theorem 3.4.1 is a special case.

[16] The compact support of the kernel function in Masry (1996) can be easily relaxed, following the line of proof in Hansen (2008, Theorem 4). Masry (1996) only allows continuous regressors, which can also be extended to the case of mixed regressors. Since X_i^d and Y_i only take finite number of possible values, they have no impact on the uniform probability order.

First, we decompose $(h!)^{1/2} D_n$ as follows:

$$\begin{aligned}
(h!)^{1/2} D_n &= (h!)^{1/2} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \sum_{i=1}^{n} \left[\widehat{F}(Z_i|X_i,r) - \widehat{F}(Z_i|X_i,s)\right]^2 \\
&= (h!)^{1/2} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \sum_{i=1}^{n} \Big\{ \left[F(Z_i|X_i,r) - F(Z_i|X_i,s)\right]^2 \\
&\quad + \left[\widehat{F}(Z_i|X_i,r) - F(Z_i|X_i,r) - \widehat{F}(Z_i|X_i,s) + F(Z_i|X_i,s)\right]^2 \\
&\quad + 2\left[F(Z_i|X_i,r) - F(Z_i|X_i,s)\right] \\
&\quad \times \left[\widehat{F}(Z_i|X_i,r) - F(Z_i|X_i,r) - \widehat{F}(Z_i|X_i,s) + F(Z_i|X_i,s)\right]\Big\} \\
&\equiv D_{n1} + D_{n2} + 2D_{n3}.
\end{aligned}$$

Under $H_1(n^{-1/2}(h!)^{-1/4})$, we prove the theorem by showing that (i) $D_{n1} \xrightarrow{P} \mu_0$, (ii) $D_{n2} - B_n \xrightarrow{d} N\left(0, \sigma_0^2\right)$, (iii) $D_{n3} = o_P(1)$, (iv) $\widehat{B}_n = B_n + o_P(1)$, and (v) $\widehat{\sigma}_n^2 = \sigma_0^2 + o_P(1)$. For (i), $D_{n1} = n^{-1} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \sum_{i=1}^{n} \delta_{n,rs}(\varsigma_i)^2 = \mu_0 + o_P(1)$ under $H_1(n^{-1/2}(h!)^{-1/4})$. It remains to show (ii)-(iv).

To show (ii), we first apply Lemma 3.8.1 to obtain

$$\begin{aligned}
D_{n2} &= (h!)^{1/2} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \sum_{i=1}^{n} \left[\widehat{F}(Z_i|X_i,r) - F(Z_i|X_i,r) - \widehat{F}(Z_i|X_i,s) + F(Z_i|X_i,s)\right]^2 \\
&= (h!)^{1/2} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \sum_{i=1}^{n} \Big[e_1'[\overline{\mathbf{S}}_r(X_i)]^{-1}\mathbf{V}_{nr}(\varsigma_i) - e_1'[\overline{\mathbf{S}}_s(X_i)]^{-1}\mathbf{V}_{ns}(\varsigma_i) \\
&\quad + e_1'[\overline{\mathbf{S}}_r(X_i)]^{-1}\overline{\mathbf{B}}_r(\varsigma_i) - e_1'[\overline{\mathbf{S}}_s(X_i)]^{-1}\overline{\mathbf{B}}_s(\varsigma_i) + O_P(\nu_n^2 + \nu_n(\|h\|^2 + \|\lambda\|))\Big]^2 \\
&= (h!)^{1/2} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \sum_{i=1}^{n} \left[e_1'\left\{[\overline{\mathbf{S}}_r(X_i)]^{-1}\mathbf{V}_{nr}(\varsigma_i) - [\overline{\mathbf{S}}_s(X_i)]^{-1}\mathbf{V}_{ns}(\varsigma_i)\right\}\right]^2 \\
&\quad + 2(h!)^{1/2} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \sum_{i=1}^{n} e_1'\left\{[\overline{\mathbf{S}}_r(X_i)]^{-1}\mathbf{V}_{nr}(\varsigma_i) - [\overline{\mathbf{S}}_s(X_i)]^{-1}\mathbf{V}_{ns}(\varsigma_i)\right\} \\
&\qquad \times e_1'\left\{[\overline{\mathbf{S}}_r(X_i)]^{-1}\overline{\mathbf{B}}_r(\varsigma_i) - [\overline{\mathbf{S}}_s(X_i)]^{-1}\overline{\mathbf{B}}_s(\varsigma_i)\right\} \\
&\quad + (h!)^{1/2} \sum_{i=1}^{n} \left[e_1'\left\{[\overline{\mathbf{S}}_r(X_i)]^{-1}\overline{\mathbf{B}}_r(\varsigma_i) - [\overline{\mathbf{S}}_s(X_i)]^{-1}\overline{\mathbf{B}}_s(\varsigma_i)\right\}\right]^2 \\
&\quad + n(h!)^{1/2} O_P\left(\nu_n^2 + \nu_n(\|h\|^2 + \|\lambda\|)\right) O_P\left(\nu_n + \|h\|^2 + \|\lambda\|\right) \\
&\equiv D_{n21} + 2D_{n22} + D_{n23} + o_P(1) \qquad (8.3)
\end{aligned}$$

where the definitions of D_{n21}, D_{n22}, and D_{n23} are self-evident. Using the notation defined above eq. (4.9), we have $D_{n21} = \frac{(h!)^{1/2}}{(n-1)^2} \sum_{i=1}^{n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} [\sum_{j=1}^{n} \varphi_{rs}\ (\xi_i, \xi_j)]^2$, where $\varphi_{rs}(\xi_i, \xi_j) \equiv \mathbf{K}_r(\zeta_j; X_i)\, \overline{\mathbf{1}}_{Z_i,r}(\varsigma_j) - \mathbf{K}_s(\zeta_j; X_i)\, \overline{\mathbf{1}}_{Z_i,s}(\varsigma_j)$. Decompose D_{n21} as follows

$$\begin{aligned}
D_{n21} &= \frac{(h!)^{1/2}}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \varphi_{rs}(\xi_i, \xi_j)\, \varphi_{rs}(\xi_i, \xi_k) \\
&= \frac{(h!)^{1/2}}{n^2} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \sum_{k\neq i,j}^{n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \varphi_{rs}(\xi_i, \xi_j)\, \varphi_{rs}(\xi_i, \xi_k) \\
&\quad + \frac{(h!)^{1/2}}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \varphi_{rs}(\xi_i, \xi_j)^2 \\
&\quad + \frac{2\,(h!)^{1/2}}{n^2} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \varphi_{rs}(\xi_i, \xi_j)\, \varphi_{rs}(\xi_i, \xi_i) \\
&\equiv V_n + B_n + R_n, \text{ say.}
\end{aligned} \tag{8.4}$$

Let $\overline{\varphi}_{rs}(\xi_i, \xi_j, \xi_k) \equiv [\varphi_{rs}(\xi_i, \xi_j)\, \varphi_{rs}(\xi_i, \xi_k) + \varphi_{rs}(\xi_j, \xi_i)\, \varphi_{rs}(\xi_j, \xi_k) + \varphi_{rs}(\xi_k, \xi_i)\, \varphi_{rs}(\xi_k, \xi_j)]/3$. Then

$$V_n = \frac{6\,(h!)^{1/2}}{n^2} \sum_{1\leq i<j<k\leq n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \overline{\varphi}_{rs}(\xi_i, \xi_j, \xi_k) = \frac{(n-1)(n-2)}{n} \overline{V}_n,$$

where $\overline{V}_n \equiv \frac{6(h!)^{1/2}}{n(n-1)(n-2)} \sum_{1\leq i<j<k\leq n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \overline{\varphi}_{rs}(\xi_i, \xi_j, \xi_k)$. Note that for all $i \neq j \neq k$, $\theta \equiv E[\overline{\varphi}_{rs}(\xi_i, \xi_j, \xi_k)] = 0$, $\overline{\varphi}_{rs,1}(a) \equiv E[\overline{\varphi}_{rs}(a, \xi_j, \xi_k)] = 0$, and $\overline{\varphi}_{rs,2}(a, \widetilde{a}) \equiv E[\overline{\varphi}_{rs}(a, \widetilde{a}, \xi_k)] = \frac{1}{3} E[\varphi_{rs}(\xi_k, a)\ \varphi_{rs}(\xi_k, \widetilde{a})]$. Let $\overline{\varphi}_{rs,3}(a, \widetilde{a}, \overline{a}) \equiv \overline{\varphi}_{rs}(a, \widetilde{a}, a) - \overline{\varphi}_{rs,2}(a, \widetilde{a}) - \overline{\varphi}_{rs,2}(a, \overline{a}) - \overline{\varphi}_{rs,2}(\widetilde{a}, a)$. By the Hoeffding decomposition,

$$\overline{V}_n = 3H_n^{(2)} + H_n^{(3)},$$

where $H_n^{(2)} \equiv \frac{2(h!)^{1/2}}{n(n-1)} \sum_{1\leq i<j\leq n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \overline{\varphi}_{rs,2}(\xi_i, \xi_j)$ and $H_n^{(3)} \equiv \frac{6(h!)^{1/2}}{n(n-1)(n-2)} \sum_{1\leq i<j<k\leq n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \overline{\varphi}_{rs,3}(\xi_i, \xi_j, \xi_k)$. Noting that $E[\overline{\varphi}_{rs,3}(a, \widetilde{a}, \xi_i)] = 0$ and that $\overline{\varphi}_{rs,3}$ is symmetric in its arguments by construction, it is straightforward to show that $E[H_n^{(3)}] = 0$ and $E[H_n^{(3)}]^2 = O(n^{-3}(h!)^{-1})$. Hence, $H_n^{(3)} = O_P(n^{-3/2}(h!)^{-1/2}) = o_P(n^{-1})$ by the Chebyshev inequality. It follows that $V_n = \frac{n(n-2)}{n-1} \overline{V}_n = \{1 + o(1)\}\, \mathcal{H}_n + o_P(1)$, where

$$\mathcal{H}_n \equiv \frac{2\,(h!)^{1/2}}{n} \sum_{1\leq i\leq j\leq n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} 3\overline{\varphi}_{rs,2}(\xi_i, \xi_j)$$

$$= \frac{2(h!)^{1/2}}{n} \sum_{1 \le i < j \le n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \int \varphi_{rs}(a, \xi_i) \varphi_{rs}(a, \xi_j) F_\xi(da).$$

As $\mathcal{H}_n$ is a second order degenerate U-statistic, it is straightforward but tedious to verify that all the conditions of Theorem 1 of Hall (1984) are satisfied, implying that a central limit theorem applies to $\mathcal{H}_n$: $\mathcal{H}_n \stackrel{d}{\to} N(0, \sigma_0^2)$, where the asymptotic variance of $\mathcal{H}_n$ is given by $\sigma_0^2 \equiv \lim_{n\to\infty} \sigma_n^2$ and $\sigma_n^2 \equiv 2h! E_i E_j [\sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \int \varphi_{rs}(\xi, \xi_i) \varphi_{rs}(\xi, \xi_j) F_\xi(d\xi)]^2 = 2h! E_i E_j [\sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \int [\mathbf{K}_r(\zeta_i; x) \overline{\mathbf{1}}_{z,r}(\varsigma_i) - \mathbf{K}_s(\zeta_i; x) \overline{\mathbf{1}}_{z,s}(\varsigma_i)] [\mathbf{K}_r(\zeta_j; x) \overline{\mathbf{1}}_{z,r}(\varsigma_j) - \mathbf{K}_s(\zeta_j; x,) \overline{\mathbf{1}}_{z,s}(\varsigma_j)] F_\xi(d\xi)]^2$. Consequently

$$V_n \stackrel{d}{\to} N(0, \sigma_0^2). \tag{8.5}$$

For R_n, it is easy to verify that $E(R_n) = 0$ and $E(R_n^2) = O(n(h!)^{-1}) = o(1)$. So $R_n = o_P(1)$ by the Chebyshev inequality. Combined with (8.4) and (8.5), we have

$$D_{n21} - B_n \stackrel{d}{\to} N(0, \sigma_0^2). \tag{8.6}$$

Let $b_{rs}(\varsigma_i) \equiv e_1' \{[\overline{\mathbf{S}}_r(X_i)]^{-1} \overline{\mathbf{B}}_r(\varsigma_i) - [\overline{\mathbf{S}}_s(X_i)]^{-1} \overline{\mathbf{B}}_s(\varsigma_i)\}$. Then $D_{n22} = \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} (D_{n22,rs1} - D_{n22,rs2})$, where $D_{n22,rs1} \equiv (h!)^{1/2} \sum_{i=1}^n e_1' [\overline{\mathbf{S}}_r(X_i)]^{-1} \mathbf{V}_{nr}(\varsigma_i) b_{rs}(\varsigma_i)$ and $D_{n22,rs2} \equiv (h!)^{1/2} \sum_{i=1}^n e_1' [\overline{\mathbf{S}}_s(X_i)]^{-1} \mathbf{V}_{ns}(\varsigma_i) b_{rs}(\varsigma_i)$. Write

$$\begin{aligned} D_{n22,rs1} &= n^{-1}(h!)^{1/2} \sum_{i=1}^n \sum_{j \ne i}^n e_1' [\overline{\mathbf{S}}_r(X_i)]^{-1} K_{h\lambda}(X_j, X_i) \mathbf{1}_j^r \boldsymbol{\tau}_h(X_j^c - X_i^c) \overline{\mathbf{1}}_{Z_i,r}(\varsigma_j) b_{rs}(\varsigma_i) \\ &\quad + n^{-1}(h!)^{1/2} \sum_{i=1}^n e_1' [\overline{\mathbf{S}}_r(X_i)]^{-1} K_{h\lambda}(X_i, X_i) \mathbf{1}_i^r \boldsymbol{\tau}_h(X_i^c - X_i^c) \overline{\mathbf{1}}_{Z_i,r}(\varsigma_i) b_{rs}(\varsigma_i) \\ &\equiv D_{n22,rs1a} + D_{n22,rs1b}, \text{ say.} \end{aligned}$$

Noting that $b_{rs}(\varsigma_i) = O_P(\|h\|^2 + \|\lambda\|)$, it is straightforward to show that $D_{n22,rs1b} = O_P((h!)^{-1/2} (\|h\|^2 + \|\lambda\|)) = o_P(1)$. Noting that $E(D_{n22,rs1a}) = 0$ and $E(D_{n22,rs1a}^2) = O(nh!(\|h\|^2 + \|\lambda\|)^2) = o(1)$, we have $D_{n22,rs1a} = o_P(1)$ by the Chebyshev inequality. Similarly, we can show that $D_{n22,rs1b} = o_P(1)$ and thus $D_{n22,rs1} = o_P(1)$. By the same token, $D_{n22,rs2} = o_P(1)$. It follows that

$$D_{n22} = o_P(1). \tag{8.7}$$

By Lemma 3.8.1 and Assumption A.5, we have $D_{n23} = n(h!)^{1/2} O_P(\|h\|^4) = O_P(n\|h\|^4 (h!)^{1/2}) = o_P(1)$. This, in conjunction with (8.3), (8.6) and (8.7), implies that $D_{n2} - B_n \stackrel{d}{\to} N(0, \sigma_0^2)$.

Next, we show (iii). By Lemma 3.8.1, under $H_1\left(n^{-1/2}(h!)^{-1/4}\right)$ we have

$$
\begin{aligned}
D_{n3} &= \sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1} n^{-1/2}(h!)^{1/4}\sum_{i=1}^{n}\left[e_1'[\overline{\mathbf{S}}_r(X_i)]^{-1}\mathbf{V}_{nr}(\varsigma_i) - e_1'[\overline{\mathbf{S}}_s(X_i)]^{-1}\mathbf{V}_{ns}(\varsigma_i)\right) \\
&\quad + e_1'[\overline{\mathbf{S}}_r(X_i)]^{-1}\overline{\mathbf{B}}_r(\varsigma_i) - e_1'[\overline{\mathbf{S}}_s(\zeta_i)]^{-1}\overline{\mathbf{B}}_s(\varsigma_i)\big]\,\delta_{n,rs}(\varsigma_i) \\
&\quad + n^{1/2}(h!)^{1/4}O_P\left(\nu_n^2 + \nu_n\left(\|h\|^2 + \|\lambda\|\right)\right) \\
&\equiv \sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1}\left[D_{n31,rs} - D_{n32,rs} + D_{n33,rs} - D_{n34,rs}\right] + o_P(1),
\end{aligned}
$$

where, for example, $D_{n31,rs} \equiv n^{-1/2}(h!)^{1/4}\sum_{i=1}^{n} e_1'[\overline{\mathbf{S}}_r(X_i)]^{-1}\ \mathbf{V}_{nr}(\varsigma_i)\,\delta_{n,rs}(\varsigma_i)$, and $D_{n3l,rs}$, $l = 2, 3, 4$, are analogously defined. Decompose

$$
\begin{aligned}
D_{n31,rs} &= n^{-3/2}(h!)^{1/4}\sum_{i=1}^{n}\sum_{j\neq i}^{n} e_1'[\overline{\mathbf{S}}_r(X_i)]^{-1}\boldsymbol{\tau}_h\left(X_j^c - X_i^c\right)K_{h\lambda}(X_j, X_i)\,\overline{\mathbf{1}}_j^r\overline{\mathbf{1}}_{Z_i,r}(\varsigma_j)\,\delta_{n,rs}(\varsigma_i) \\
&\quad + n^{-3/2}(h!)^{1/4}\sum_{i=1}^{n} e_1'[\overline{\mathbf{S}}_r(X_i)]^{-1}\boldsymbol{\tau}_h(0)\,K_{h\lambda}(X_i, X_i)\,\overline{\mathbf{1}}_i^r\overline{\mathbf{1}}_{Z_i,r}(\varsigma_i)\,\delta_{n,rs}(\varsigma_i) \\
&\equiv D_{n31,rs1} + D_{n31,rs2},\ \text{say.}
\end{aligned}
$$

It is easy to show that $D_{n31,rs2} = O_P(n^{-1/2}(h!)^{-3/4}) = o_P(1)$ by Assumption A.5. For $D_{n31,rs1}$, noting that $E[D_{n31,rs1}] = 0$ and

$$
\begin{aligned}
&E\left[D_{n31,rs1}\right]^2 \\
&= n^{-3}(h!)^{1/2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{j\neq i,i'}^{n} E\left\{e_1'[\overline{\mathbf{S}}_r(X_i)]^{-1}\boldsymbol{\tau}_h\left(X_j^c - X_i^c\right)K_{h\lambda}(X_j, X_i)\,\mathbf{1}_j^r\overline{\mathbf{1}}_{Z_i,r}(\varsigma_j)\,\delta_{n,rs}(\varsigma_i)\right. \\
&\quad \left.\times e_1'[\overline{\mathbf{S}}_r(X_{i'})]^{-1}\boldsymbol{\tau}_h\left(X_j^c - X_{i'}^c\right)K_{h\lambda}(X_j, X_{i'})\,\overline{\mathbf{1}}_{Z_{i'},r}(\varsigma_j)\,\delta_{n,rs}(\varsigma_{i'})\right\} \\
&\quad + n^{-3}(h!)^{1/2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} E\left\{e_1'[\overline{\mathbf{S}}_r(X_i)]^{-1}\boldsymbol{\tau}_h\left(X_j^c - X_i^c\right)K_{h\lambda}(X_j, X_i)\,\mathbf{1}_j^r\overline{\mathbf{1}}_{Z_i,r}(\varsigma_j)\,\delta_{n,rs}(\varsigma_i)\right. \\
&\quad \left.\times e_1'[\overline{\mathbf{S}}_r(X_j)]^{-1}\boldsymbol{\tau}_h\left(X_i^c - X_j^c\right)K_{h\lambda}(X_i, X_j)\,\mathbf{1}_i^r\overline{\mathbf{1}}_{Z_j,r}(\varsigma_i)\,\delta_{n,rs}(\varsigma_j)\right\} \\
&= O\left((h!)^{1/2} + n^{-1}(h!)^{-1/2}\right) = o(1),
\end{aligned}
$$

we have $D_{n31,rs1} = o_P(1)$ by the Chebyshev inequality. Hence $D_{n31,rs} = o_P(1)$. Similarly $D_{n32,rs} = o_P(1)$. Noting that $\sup_\varsigma\left|\overline{\mathbf{B}}_r(\varsigma)\right| = O(\|h\|^2 + \|\lambda\|)$, we have

$$
D_{n33,rs} \leq n^{1/2}(h!)^{1/4}O\left(\|h\|^2 + \|\lambda\|\right)n^{-1}\sum_{i=1}^{n}|\delta_{n,rs}(\varsigma_i)| = O_P\left(n^{1/2}\|h\|^2(h!)^{1/4}\right) = o_P(1).
$$

Similarly $D_{n34,rs} = o_P(1)$. Consequently, $D_{n3} = o_P(1)$.

We now show (iv). Noting that $a^2 - b^2 = (a-b)^2 + 2(a-b)b$, we have $\widehat{B}_n - B_n = \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} B_{n1,rs} + 2\sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} B_{n2,rs}$, where

$$
\begin{aligned}
B_{n1,rs} &\equiv \frac{(h!)^{1/2}}{n^2} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \{\widehat{\alpha}_{ij,r} - \widehat{\alpha}_{ij,s}\}^2, \\
B_{n2,rs} &\equiv \frac{(h!)^{1/2}}{n^2} \sum_{i=1}^{n} \sum_{j\neq i}^{n} [\widehat{\alpha}_{ij,r} - \widehat{\alpha}_{ij,s}][\mathbf{K}_r\left(\zeta_j; X_i\right) \overline{\mathbf{1}}_{Z_i,r}(\varsigma_j) - \mathbf{K}_s\left(\zeta_j; X_i,\right) \overline{\mathbf{1}}_{Z_i,s}(\varsigma_j)],
\end{aligned}
$$

and $\widehat{\alpha}_{ij,r} = e_1'[\mathbf{S}_{nr}(X_i)]^{-1} \boldsymbol{\tau}_h \left(X_j^c - X_i^c\right) K_{h\lambda}(X_j, X_i) \mathbf{1}_j^r \widehat{\mathbf{1}}_{Z_i,r}(\varsigma_j) - \mathbf{K}_r\left(\zeta_j, X_i\right) \overline{\mathbf{1}}_{Z_i,r}(\varsigma_j)$. Noting that $[\mathbf{S}_{nr}(X_i)]^{-1} = [\overline{\mathbf{S}}_r(X_i)]^{-1} + O_P(\nu_n)$ and $\widehat{\mathbf{1}}_{z,r}(\varsigma_j) - \overline{\mathbf{1}}_{z,r}(\varsigma_j) = F(z|X_j, r) - \widehat{F}(z|X_j, r) = O_P(\nu_n + \|h\|^2 + \|\lambda\|)$ uniformly in X_j and z, we have
$\widehat{\alpha}_{ij,r} = e_1'[\overline{\mathbf{S}}_r(X_i)]^{-1} \boldsymbol{\tau}_h \left(X_j^c - X_i^c\right) K_{h\lambda}(X_j, X_i) \times \mathbf{1}_j^r \{\widehat{\mathbf{1}}_{Z_i,r}(\varsigma_j) - \overline{\mathbf{1}}_{Z_i,r}(\varsigma_j)\} + O_P(\nu_n)$. It follows that

$$
\begin{aligned}
|B_{n1,rs}| &\leq \frac{(h!)^{1/2}}{(n-1)^2} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \left\|\boldsymbol{\tau}_h\left(X_j^c - X_i^c\right) K_{h\lambda}(X_j, X_i)\right\|^2 \times O_P\left(\left(\nu_n + \|h\|^2 + \|\lambda\|\right)^2\right) \\
&= O_P\left((h!)^{-1/2}\left(\nu_n^2 + \|h\|^4 + \|\lambda\|^2\right)\right) = o_P(1),
\end{aligned}
$$

and similarly $|B_{n2,rs}| = O_P\left((h!)^{-1/2}\left(\nu_n + \|h\|^2 + \|\lambda\|\right)\right) = o_P(1)$ under Assumption A.5. Consequently, $\widehat{B}_n - B_n = o_P(1)$.

For (v), noticing that

$$
\begin{aligned}
\widehat{\beta}_{ij,rs} &= \frac{1}{n} \sum_{l=1}^{n} \left\{\mathbf{K}_r(\zeta_i, X_l) \overline{\mathbf{1}}_{Z_l,r}(\varsigma_i) - \mathbf{K}_s(\zeta_i, X_l) \overline{\mathbf{1}}_{Z_l,s}(\varsigma_i)\right\} \\
&\quad \times \left\{\mathbf{K}_r\left(\zeta_j, X_l\right) \overline{\mathbf{1}}_{Z_l,r}(\varsigma_j) - \mathbf{K}_s\left(\zeta_j, X_l\right) \overline{\mathbf{1}}_{Z_l,s}(\varsigma_j)\right\} + o_P(1) \\
&= \int \left\{\mathbf{K}_r(\zeta_i, x) \overline{\mathbf{1}}_{z,r}(\varsigma_i) - \mathbf{K}_s(\zeta_i, x) \overline{\mathbf{1}}_{z,s}(\varsigma_i)\right\} \{\mathbf{K}_r\left(\zeta_j, x\right) \overline{\mathbf{1}}_{z,r}(\varsigma_j) \\
&\quad - \mathbf{K}_s\left(\zeta_j, x\right) \overline{\mathbf{1}}_{z,s}(\varsigma_j)\} F_\xi(d\xi) + o_P(1),
\end{aligned}
$$

we have $\widehat{\sigma}_n^2 = \sigma_0^2 + o_p(1)$ by the law of large numbers for U-statistics. ■

Proof of Theorems 3.4.3

Using the notation defined in the proof of Theorem 3.4.2, we again write $n^{-1} D_n = n^{-1}(h!)^{-1/2}(D_{n1} + D_{n2} + 2D_{n3})$. Under H_1, it is easy to show that $n^{-1}(h!)^{-1/2} D_{n1} = \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} E[F(Z_i|X_i, r) - F(Z_i|X_i, s)]^2 + o_P(1)$, $n^{-1}(h!)^{-1/2} D_{n2} = O_P(\nu_n^2 + \|h\|^4 +$

$\|\lambda\|^2) = o_P(1)$, and $n^{-1}(h!)^{-1/2} D_{n3} = O_P(\nu_n + \|h\|^2 + \|\lambda\|) = o_P(1)$. On the other hand, $n^{-1}(h!)^{-1/2}\widehat{B}_n = O_P(n^{-1}) = o_P(1)$ and $\widehat{\sigma}_n^2 = \sigma_0^2 + o_P(1)$. It follows that $n^{-1}(h!)^{-1/2} T_n = (n^{-1}D_n - n^{-1}(h!)^{-1/2}\widehat{B}_n)/\sqrt{\widehat{\sigma}_n^2} \xrightarrow{P} \sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1} E[F(Z_i|X_i, r) - F(Z_i|X_i, s)]^2/\sigma_0$, and the conclusion follows. ■

3.9 References

Aitchison, J., Aitken, C. G. G., 1976. Multivariate binary discrimination by the kernel method. Biometrika 63, 413-420.

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

Boente, G., Fraiman, R., 1991. Strong uniform convergence rates for some robust equivariant nonparametric regression estimates for mixing processes. International Statistical Review 59, 355-372.

Chiappori, P.-A., Heckman, J. J., 1999. Testing for moral hazard using dynamic data. Manuscript, Dept. of Economics, Univ. of Chicago.

Chiappori, P.-A., Salanié, B. 1997. Empirical contract theory: the case of insurance data. European Economic Review 41, 943-950.

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

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., 2005. Asymmetric information and learning: evidence from the automobile insurance market. The Review of Economics and Statistics 87, 197-207.

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

Crocker, K.-J., Snow, A., 2000. The theory of risk classification. In: G. Dionne, editor, Handbook of Insurance, Chapter 8, 245-276. Kluwer Academic Publishers.

Delgado, M. A., González-Manteiga, W., 2001. Significance testing in nonparametric regression based on the bootstrap. Annals of Statistics 29, 1469-1507.

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

Dionne, G., Gourieroux C., 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., Michaud, P.-C., Dahchour, M., 2004. Separating moral hazard from adverse selection in automobile insurance: Longitudinal evidence from France. Cahiers de recherche 0420, CIRPEE.

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

Hall, P., 1984. Central limit theorem for integrated square error of multivariate nonparametric density estimators. Journal of Multivariate Analysis 14, 1-16.

Hansen, B. E., 2008. Uniform convergence rates for kernel estimation with dependent data. Econometric Theory 24, 726-748.

Huang, M., White, H., 2009. A flexible test for conditional independence. Working Paper, Dept. of Economics, UCSD.

Huang, T.-M., 2009. Testing conditional independence using maximal nonlinear conditional independence. Working Paper, National Chengchi University.

Jullien, B., Salanié, B., Salanié, F., 2007. Screening risk averse agents under moral hazard. Economic Theory 30, 151-169.

Kim, H., Kim, D., Im, S., Hardin, J. W., 2009. Evidence of asymmetric information in the automobile insurance market: dichotomous versus multinomial measurement of insurance coverage. The Journal of Risk and Insurance 76, 343-366.

Li, Q., Racine, J., 2007. Nonparametric Econometrics: Theory and Practice. Princeton University Press, Princeton and Oxford.

Li, Q., Racine, J., 2008. Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. Journal of Business and Economic Statistics 26, 423-434.

Masry, E., 1996. Multivariate local polynomial regression for time series: uniform strong consistency rates. Journal of Time Series Analysis 17, 571-599.

Paparoditis, E., Politis, D. N., 2000. The local bootstrap for kernel estimators under general dependence conditions. Annals of the Institute of Statistical Mathematics 52, 139-159.

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

Racine, J., Li, Q., 2004. Nonparametric estimation of regression functions with both categorical and continuous data, Journal of Econometrics 119, 99-130.

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.

Robinson, P. M., 1991. Consistent nonparametric entropy-based testing. Review of Economic Studies 58, 437-453.

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.

Saito, K., 2009. Testing for asymmetric information in the automobile insurance market under rate regulation. The Journal of Risk and Insurance 73, 335-356.

Song, K., 2009. Testing conditional independence via Rosenblatt transform. Annals of Statistics 37, 4011-4045.

Su, L., White, H., 2007. A consistent characteristic-function-based test for conditional independence. Journal of Econometrics 141, 807-834.

Su, L., White, H., 2008. A nonparametric Hellinger metric test for conditional independence. Econometric Theory 24, 829-864.

Su, L., White, H., 2010. Local polynomial quantile regression for dependent heterogeneous processes: uniform Bahadur representation with applications to testing conditional independence. Working Paper, Dept. of Economics, UCSD.

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

Chapter 4

Asymmetric Information in the Accident Insurance*

4.1 Introduction

Over the last years testing for asymmetric information in insurance markets has gained much popularity. This leads to narrowing the gap between theory and empirical evidence. Empirical results also show directions for further theoretical developments.

The theory of asymmetric information is well understood for a long time. The phenomenon of "adverse selection"was first analyzed by Akerlof (1970). Rothschild and Stiglitz (1976) introduced the notion of adverse selection in insurance markets. Moral hazard, the second constituent of asymmetric information, has been developed by Holmström (1979) and with application to insurance markets by Shavell (1979). The models for both phenomena predict a positive correlation between risk and coverage. Although it is in general difficult to disentangle adverse selection and moral hazard, tests for asymmetric information as a whole are possible. While theory has been highly developed, empirical studies lagged behind. One reason is the scare availability of data sets in this field. A cornerstone in this development was Chiappori and Salanié (2000). They introduced bivariate testing procedures for the positive correlation property which became the standard test in this field. Google scholar counts approximately 500 citations on their paper. The basic idea is to define binary proxy variables for risk, usually measured as ex post risk, and for coverage, i.e., the chosen contract. Then tests for conditional correlation (parametric ones and a simple nonparametric one) are conducted. Their procedure was applied amongst others in Cohen (2005), Saito (2006), Aarbu (2010), Spindler et al. (2011), Muermann and Straka (2011).

*This chapter is based on joint work with Liangjun Su.

In a recent paper Su and Spindler (2011) introduce a general nonparametric test for asymmetric information. The basic idea is an alternative resp. more general definition of what asymmetric information means. The absence of asymmetric information means that the choice of a contract Y provides no information for predicting the "performance" variable Z (e.g., the number of claims or the sum of reimbursements), conditional on the vector X of all exogenous variables. Therefore one can transform the problem of testing the absence of asymmetric information into a test for conditional independence: $F(Z|X,Y) = F(Z|X)$ where, $F(Z|X,Y)$ denotes the conditional cumulative distribution function (CDF) of Z given (X,Y).
The advantage of nonparametric tests is that they do not rely on certain distributional and functional specifications. For example probit models rely on the normal assumption which might be questionable in insurance applications. Moreover, insurance companies use non-linear influences, e.g., cross effects, for risk classification. These might be unknown to the econometrician despite his access to the data, as the pricing formula is usually not disclosed. The main focus of this paper is to apply both parametric and nonparametric tests to a novel data set and to compare the results.

The rest of the paper is structured as follows. In section 2 we give an introduction to accident insurance. Section 3 presents the data set we use. Section 4 gives a repetition of the applied testing procedures. In section 5 we show the results and finally conclude in section 6.

4.2 The Accident Insurance

The accident insurance is very elementary for many insurees as it covers risks which touch the existence of individuals. In Germany there are two pillars concerning the accident insurance: a compulsory and a private voluntary accident insurance. The statutory accident insurance covers only risks related to the workplace. These are risks like working and commuting accidents and occupational diseases. In case of such an event this insurance covers the costs for reconstituting the health or pays a pension in the case of incapacity for work. As the compulsory accident insurance applies only to employees (e.g., not for freelancers) and the indemnification payments from the compulsory accident insurance are very limited in the case of an accident and - as mentioned before - only accidents related to work are covered, a private accident insurance is a very important supplement.

The (private) accident insurance usually covers the following eventualities: invalidity and dismemberment, death, and a hospital per diem[1] in the case of a hospital stay. These are the

[1] Strictly speaking, the daily payment in the case of a hospital stay consists of two parts, a hospital per

basic risks which are covered by default. Besides, some insurance companies offer additional insurance payments, e.g., reimbursements of costs for treatment at a health resort or costs for plastic surgery. It is important to mention that the eventualities are only covered if they arise as a consequence of an accident.
For each eventuality the insuree can choose the level of payments in case of occurrence of the event insured. The premium depends on these chosen levels. Additionally, the insurance companies can use variables like occupation, sex, age and so on for risk classification and rating. These observables also determine the insurance premium. In case of invalidity the insuree chooses the level (i.e., insured sum) 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 invalidity case the insuree can choose between certain schedules of progression. For example the insuree can choose to get the full amount for total disability already from disabilities of 50 % onwards or to get from certain level of disability (in most cases relatively high degrees of disability) on twice or five times the chosen basic amount. The chosen progression clearly influences the insurance premium paid by the insuree.

To close this section some facts and figures about the accident insurance in Germany (GDV (2010)) are given: In Germany 40.8 % of the households have an accident insurance. In the year 2009 the premium income was 6,389 million Euro, claims expenditure 2,928 million Euro. This results in a claims ratio of 58.2 %.

4.3 The Data Set

For our analysis we have access to a 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 and therefore in the data set. 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.

diem and a convalescence allowance, but this distinction is only of minor importance for the analysis.

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

Under one contract several different persons can be insured. For example a father can take out a policy for him, his wife and his kids. Such constellations are also captured in the database. For each insured risk personal data and the claim history are filed. As we want to test the risk assessment of the individual for itself, not the ability to assess the risk of other individuals, we restrict to contracts containing only one insured person.
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.

When 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 set “historic”and is not in the sample. Therefore there is a sample selection with a tendency for “good”risks being in the sample given the same date of signing (“attrition bias”).
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 has still a size of $n = 77,125$. It is reasonable that the individuals might have an information advantage concerning their accident probability but not concerning the timing of an accident.[2] Therefore this procedure should not influence the results or distort them.

[2] According to personal communication with actuaries, the accident probability rises with age, i.e., in higher age accidents occur more often, but this increase concerns good and bad risks comparably.

Table 4.1: 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
insurance sum for rent	0	165	1075
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 $n = 957,506$

Table 4.2: Number of claims

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

Table 4.3: Descriptive statistics of the subsample

Variable	Minimum	Mean	Maximum
age	14.6	40.3	97
insurance sum for disability	3,835	103,823	1,032,500
insurance sum for death	767	5,415	51,130
insurance sum for hospital per diem	2.5	8.3	65
cost of claims for disability	0	13.9	115,000
cost of claims for death	0	1.29	21000
cost of claims for hospital per diem	0	1.9	5,458

sample size $n = 77,125$

4.4 Methods

In this section we give a short review of the methods applied and customize them to the accident insurance. In the literature mainly parametric procedures are used. We additionally present and apply a nonparametric test which was introduced in Su and Spindler (2011).
In the following, X denotes the exogenous variables which are used for risk classification by the insurance company. Y denotes the chosen contract. In accident insurance individuals can only choose the insured sum for each eventuality, i.e., the sum which is paid out in the case of death, the sum for a 100 % invalidity (together with the chosen progression resp. factor) or the daily payment for a hospital stay, all other conditions being the same. Z measures the risk. The risk is measured as “ex post risk”, e.g., by the number of accidents or the caused damage payments by the insuree. An index i refers to a certain individual or contract which is omitted if there is no confusion.

4.4.1 Parametric Methods

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. These procedures were applied for the first time for testing asymmetric information in Chiappori and Salanié (2000) and since then have become the standard procedures in this field.

4.4.1.1 Two Probits

One approach is to define two probit models, one for the choice of the coverage Y_i (either low or high insured sum) and the other for the occurrence of an accident Z_i:

$$\begin{cases} Y_i = \mathbf{1}(X_i\beta + \varepsilon_i > 0) \\ Z_i = \mathbf{1}(X_i\gamma + \eta_i > 0) \end{cases} \tag{4.1}$$

where ε_i and η_i are independent standard normal errors, and β and γ are coefficient vectors (as columns). Y_i is defined as 1, if the insured sum is above a predefined cut off value and 0 if equal or below. Z_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, both probit models are estimated independently and then the generalized residuals[3] $\hat{\varepsilon}_i$ and $\hat{\eta}_i$ are calculated. These are required for the following test statistic

$$W_n = \frac{\left(\sum_{i=1}^{n} \hat{\varepsilon}_i \hat{\eta}_i\right)^2}{\sum_{i=1}^{n} \hat{\varepsilon}_i^2 \hat{\eta}_i^2}. \tag{4.2}$$

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 seldom. But for coverage Y_i one must choose a cut off point to define the low and high insured sums. In order to check for robustness of our results we will choose several cut off points to see how the definition of the binary variable Y_i influences the results.

4.4.1.2 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- oder LR-test can be used.

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

4.4.2 Nonparametric Methods

4.4.2.1 A Simplified Nonparametric Test

Motivated by the χ^2-test for independence in the statistics literature, Chiappori and Salanié (2000) propose a nonparametric test for asymmetric information by restricting all variables in X_i, Y_i, and Z_i to be binary. They choose a set of m exogenous binary variables in X_i, and construct $M \equiv 2^m$ cells in which all individuals have the same values for all variables in X_i. For each cell they set up a 2×2 contingency table generated by the binary values of Y_i and Z_i, and conduct a χ^2-test for independence. This results in M test statistics, each of them is distributed asymptotically as $\chi^2(1)$ under the null hypothesis. They aggregate these M test statistics in three ways to obtain three overall test statistics for conditional independence: one is the Kolmogorov-Smirnoff test statistic that compares the empirical distribution function of the M test statistics with the CDF of the $\chi^2(1)$ distribution; the second is to count the number of rejections for the independence test for each cell which is asymptotically distributed as binomial $B(M, \alpha)$ under the null, where α denotes the significance level of the χ^2 test within each cell; and the third is the sum of all test statistics for each individual cell, which is asymptotically $\chi^2(M)$ distributed under the null.

One drawback is that all variables have to be binary or be forced to be binary which leads to a loss of information. In the case of accident insurance all observables used for pricing are binary, but, e.g., the insuree can choose the insured sums freely and therefore Y is not restricted to be binary. If one would like to control for age or the duration of the contract this approach has only limited power.

4.4.2.2 A General Nonparametric Approach

An alternative interpretation of asymmetric information resp. its absence is that the chosen contract contains no information concerning the distribution of accidents resp. vice versa. In the case of no asymmetric information (null hypothesis) one would expect that

$$H_0 : F(Y|X, Z) = F(Y|X) \tag{4.3}$$

where F denotes the conditional distribution function, Y the chosen contract and Z the risk. This means that the number of accidents has no predictive power for the choice of contract. Su and Spindler (2011) propose a test which builds upon this principle. To apply this test, Y has to be discrete, Z can be discrete or continuous and X can consist of both discrete and continuous variables.

They estimate the conditional distribution functions by using the local linear method and permit smoothing also for discrete variables. The test statistic is given by

$$D_n \equiv \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \sum_{i=1}^{n} \left[\widehat{F}\left(Z_i|X_i, r\right) - \widehat{F}\left(Z_i|X_i, s\right) \right]^2,$$

where c_y denotes the number of categories of the variable Y. $\hat{F}$ is estimated with local linear regression and smoothing of the continuous and discrete variables in X, as mentioned above. Under the null hypothesis the test statistic is asymptotically normally distributed (after removal of a bias term) and the test is consistent under the global alternative. For further details and a Monte Carlo simulation which confirms the power of the test in small samples we refer to Su and Spindler (2011). As for nonparametric procedures the asymptotic results are in general only a bad approximation for small samples we determine the p-values via bootstrapping.

As mentioned above, the insuree can chose the insured sum for the eventualities freely. Therefore Y is continuous in nature, but the chosen sums are usually chosen either in thousands of Euro (death, invalidity) or in whole Euro (hospital per diem). Additionally, there are some focal points so that the variable Y can be regarded as discrete with a manageable number of categories. In the case of the accident insurance a higher chosen insured sum also leads to higher payments when an accident occurs. This would introduce a spurious positive correlation. Therefore we do not use the exact sum paid in case of occurrence of the event insured but we code Z as a binary indicator variable for the occurrence of at least one accident.

To construct the test statistic, one needs to choose both the kernel and in particular the bandwidth. As there is no data-driven procedure to choose the bandwidth for smoothing the continuous and discrete variables for our approach, we choose them according to a rule of thumb. As the choice of the bandwidth is crucial we use a multiplicative constant γ to vary the bandwidth in order to check for robustness.

4.5 Results

In this section we present the results. We analyze each eventuality separately as a joint analysis would be more confusing and not bring much additional benefit. Further we only present the results for the eventuality of invalidity. The danger of invalidity is the main reason why insurees buy accident insurance. The results for the hospital per diem can be found in the appendix.

In insurance economics there is also the puzzle of unused observables. This states that insurance companies often do not use all the available information for pricing (Finkelstein and Poterba (2006)). In the case of accident insurance only sex and the riskiness of the occupation, which is given by two categories, are used for pricing. Other variables like age, place of residence and so on are not used although they are available. Therefore we conduct the analysis both only on the variables used by the insurance company and on an extended set of observational variables which are available in the data set. The extended configuration contains in addition family status, occupation, age and information whether the insuree has chosen the option to increase the insured sum automatically on an annual basis (annual adjustment).

In order to make the parametric and nonparametric tests comparable we apply them to the same random subsample of size $n = 7,000$.

In order to test for asymmetric information several different parametric tests have been proposed in the literature, e.g., Dionne et al. (2001). Spindler et al. (2011) show that these tests deliver consistent results and therefore we restrict our analysis to the most popular procedures.

4.5.1 Invalidity

4.5.1.1 Parametric Procedures

In order to apply the introduced parametric procedures one has to define binary variables for "risk"and "coverage". Risk is measured as ex post risk and is defined as 1 if there was at least one accident resp. claim and 0 otherwise, i.e., if there has been no accident during the considered period. For the "choice of contract"we also define a binary variable which is set to 1 if the insured sum is above a cut off value and 0 if it is below or equal to this threshold. In order to check for robustness and to analyze if or how the asymmetric information varies with the level of the insured sum we variegate this cut off level. The contracts are identical, only the level of insured sum and therefore the insurance premium, which also depends on the risk variables, are different. In Tables 4.4 and 4.5 below the results for the basic and extended configuration are given according to different cut off values for the proxy for the choice of contract, i.e., choice of insured sum.

The positive correlation between risk and coverage is significant for the "low"and "middle"cut off levels ($10,000$ Euro, $50,000$ Euro and $100,000$ Euro) for both the basic and the extended configuration. For higher thresholds the correlation gets insignificant and remains positive for the original risk classification, while it even turns negative for the extended set of risk variables yet remaining insignificant. The strength of correlation is monotone decreasing with the level of the cut off value. Both the test relying on the two probits and the simple nonparametric test confirm these results. One important observation is that additional variables like age can reduce the level of positive correlation, which means a reduction of the extent of asymmetric information, but cannot completely eliminate it. For the “low”cut offs we observe a quite high positive correlation of approximately 0.5. This indicates that the choice between none or only a very basic coverage and a considerable amount of insurance is driven by asymmetric information. For higher values this effect fades out implying that the choice of really high insured sums is not driven by informational asymmetries.

Table 4.4: Results - basic configuration, subsample

	cut off point in Euro	10,000	50,000	100,000	200,000	300,000	400,000
bivariate	ρ	0.4963***	0.4014***	0.3447***	0.0853	0.0436	0.0653
probit	s.e.	0.0699	0.0662	0.0663	0.0706	0.1057	0.1214
	t-statistic	7.10	6.06	5.45	1.21	0.41	0.54
	LR-Test	46.40	33.78	27.35	1.42	0.17	0.28
two probits	W	30.45	24.31	15.17	0.84	0.56	0.19
simple np test	χ^2	34.10	26.49	24.49	5.16	0.55	0.61

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

Table 4.5: Results - extended configuration (sex, occupation, family, age, dynamic), subsample

	cut off point in Euro	10,000	50,000	100,000	200,000	300,000	400,000
bivariate	ρ	0.5510***	0.3410***	0.2347***	−0.0154	−0.0520	−0.0185
probit	s.e.	0.0846	0.0901	0.0820	0.0763	0.1106	0.1262
	t-statistic	6.52	3.79	2.88	−0.2	−0.47	−0.15
	LR-Test	32.56	15.34	10.36	0.05	0.22	0.02
two probits	W	27.44	20.04	11.14	0.15	0.15	0.11

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

4.5.1.2 Nonparametric Procedure

In this section we present the results of the nonparametric test introduced in Su and Spindler (2011). This test is robust against functional and distributional misspecification. We analyze a basic and extended setting of exogenous variables as described in the previous section. Again we have to define variables for risk and coverage. For risk we use the same definition, i.e., a binary variable with 1 if there was at least one accident and 0 otherwise.[4] Our test restricts the variable for the choice of contract to be discrete. This is no real limitation as there are some "focal points" of the insured sum which are chosen by default. For the variable Y we use different configurations which are given below. In configurations 1, 2 and 3, Y is defined as a binary variable with different cut offs. The thresholds are $200,000$ Euro, $100,000$ Euro, and $50,000$ Euro respectively. Configuration 4, 5 and 6 take some kind of "higher resolution" and are defined in the following way:

$$Y = \begin{cases} 0, & \text{if} \quad x \leq a_1 \\ 1, & \text{if} \quad a_1 < x \leq a_2 \\ 2, & \text{if} \quad x > a_2 \end{cases}$$

For configuration 4 we specify $a_1 = 50,000$ Euro and $a_2 = 200,000$ Euro, for 5 the corresponding thresholds are $100,000$ Euro and $250,000$ Euro and for configuration 6 we use $25,000$ Euro and $100,000$ Euro.

As the choice of the smoothing parameter is critical in nonparametric tests, we check different parameters γ and therefore bandwidths for robustness. A more detailed explanation for the choice of the smoothing parameter can be found in Su and Spindler (2011).

By comparing corresponding configurations we observe that additional information helps to dilute the asymmetric information. The p-values for the extended version are considerably higher than the p-values for the corresponding basic configuration.

Table 4.6: Bootstrap p-values for our nonparametric test with different numbers of choice levels and different variables

γ\Configurations	Basic 1	Basic 6	Ext. 1	Ext. 2	Ext. 3	Ext. 4	Ext. 5	Ext. 6
$\gamma = 0.75$	0.22	0.03	0.56	0.05	0.09	0.13	0.33	0.20
$\gamma = 1$	0.26	0.04	0.38	0.00	0.03	0.09	0.07	0.14
$\gamma = 1.25$	0.20	0.06	0.41	0.07	0.04	0.10	0.10	0.21
$\gamma = 1.5$	0.24	0.11	0.26	0.05	0.04	0.11	0.16	0.20

[4] Although the test is capable of using a continuous variable or a variable with more than two categories for the claim and although the payments in the case of an accident are filed we cannot use this detailed information because the exact level of disability is not filed and therefore we maintain the binary definition.

In line with the results of the parametric tests we find that for high cut off values (200, 000 Euro) there is no asymmetric information while for 100, 000 Euro and 50, 000 Euro we detect asymmetric information even for the extended version (configuration 1 versus 2 and 3). A very important finding is that dividing the variable for the choice of contract Y into more categories decreases the p-values and some of the p-values come close to the significance level of 10 %. This shows that also for the (nonparametric) test information aggregation might lead to a disguise of asymmetric information. This can be seen if we compare, for example, configuration 1 and 4. Both involve the cut off values 200, 000 Euro while the latter one also takes a smaller threshold into account.
Another interesting comparison is between configuration 2 and 6. In this case the "finer resolution"of the choice variable, i.e., taking more of the available information into account, leads to not rejecting the null hypothesis and no detection of asymmetric information.
These two comparisons show that the effect of information aggregation is undetermined.

4.5.2 Hospital per diem

Here we present the results for the hospital per diem. We apply the parametric procedures and the nonparametric test to the same subsample. For reasons of clarity we give only an overview of the results. The corresponding tables can be found in the appendix.
As the hospital per diem is part of the same contract as the insurance for invalidity, the same variables are available and pricing uses sex and the risk category of the occupation. Similar to invalidity, further variables are available but not used. Therefore we use several different settings for our analysis. We use the basic setting (sex and risk category) and an extended version which includes additionally family status, age, exact occupation, academic degree and dynamic, i.e., if the insured sum is raised automatically from year to year. For the parametric procedures we set different cut off points for the definition of the variable Y. The contracts only differ in the hospital per diem paid during a stay in hospital, all other conditions being equal.
For the nonparametric tests we define the variable Y as following: In configuration 1, Y is 0 if the payment per day is below or equal to 25 Euro and 1, if above. In configuration 2 the cut off value is 15 Euro and 20 Euro in configuration 3. In setting 4, Y is 0 if the daily allowance is below or equal to 15 Euro, 1 if between 15 and 30 Euro (including 30 Euro) and 2 if above 30 Euro a day.
We observe that for the low and middle ranges of the cut off values the parametric tests indicate a strong positive correlation between risk and coverage while for high values the correlation becomes even insignificant. The correlation decreases with increasing thresholds.

For the thresholds of 5, 10 and 15 Euro the correlation is quite high.
For the nonparametric procedure we see again that additional information (either for the risk classification X or the measurement of the contract choice Y) weakens the extent of asymmetric information (see either basic vs. extended configuration or configuration 1 vs. 4). In the nonparametric setting we do not find asymmetric information even for the low cut off point, i.e., setting 2. In this case, relying on too restrictive functional assumptions might lead to wrong results and therefore the use of standard methods might be problematic.

4.6 Conclusion

In this paper we apply both parametric tests and a nonparametric test to the accident insurance which has – as far as we know – never been analyzed in the literature under the perspective of asymmetric information. We analyze invalidity which is the most important risk covered by the contracts. For low and middle thresholds we find a quite strong correlation of risk and coverage which confirms the prediction of the basic equilibrium models of adverse selection and moral hazard. For the high cut off values we do not find a significant positive correlation. An interpretation of this finding might be that the motive to buy exceptional high insurance coverage is driven by other motives than an informational advantage of the insured. For low and middle choices of coverage informational asymmetries and their use seems to be very important.

Additionally, we also have a look at the hospital per diem although it is only of minor importance but nevertheless we make an important observation. While the parametric tests find a positive correlation between risk and coverage for the low cut off values, the nonparametric test for conditional independence does not reject the null hypothesis and indicates absence of asymmetric information. It seems that in this example functional and especially distributional departures from the standard assumptions are prevalent and influence the results of the corresponding tests.

In future studies it would be interesting to analyze the exact dependence structure resp. joint distribution of risk and coverage for the different eventualities in more detail. This might reveal interesting insights into the choice behavior of insureds

4.7 Appendix: Hospital per diem

Table 4.7: Results - basic configuration, subsample

	cut off point in Euro	5	10	15	20	25	30	40
bivariate probit	ρ	0.6761***	0.5214***	0.4124***	0.1905**	0.1243	0.1683	0.1971
	s.e.	0.0742	0.0606	0.0626	0.0821	0.0984	0.1094	0.1268
	t-statistic	9.11	8.61	6.59	2.32	1.26	1.54	1.56
	LR-Test	74.31	58.20	39.97	5.07	1.53	2.20	2.20
two probits	W	29.60	23.12	14.03	2.01	0.90	1.20	1.18
simple np test	χ^2	67.35	67.59	54.01	9.70	11.43	9.59	8.08

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

Table 4.8: Results - extended configuration (sex, academic degree, occupation, family, dynamic, age), subsample

	cut off point	5	10	15	20	25	30	40
bivariate probit	ρ	0.6163***	0.4717***	0.3610***	0.1534*	0.0780	0.1463	0.2076
	s.e.	0.0817	0.0669	0.0690	0.0866	0.1045	0.1155	0.1330
	t-statistic	7.54	6.85	5.23	1.77	0.76	1.27	1.56
two probits	W	22.73	18.26	11.02	1.05	0.30	0.79	0.93

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

Table 4.9: Bootstrap p-values for our nonparametric test with different numbers of choice levels and different variables

γ\Configurations	Basic 1	Basic 4	Ext. 1	Ext. 4	Ext. 2	Ext. 3
$\gamma = 0.75$	0.24	0.22	0.45	0.98	0.23	0.26
$\gamma = 1$	0.32	0.28	0.31	0.93	0.19	0.42
$\gamma = 1.25$	0.32	0.26	0.31	0.97	0.20	0.41
$\gamma = 1.5$	0.25	0.23	0.44	0.95	0.22	0.43

4.8 References

Aarbu, K.O., 2010. Asymmetric information - evidence form the home insurance market. Discussion Paper Norwegian School of Economics and Business Administration.

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

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

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., 2005. Asymmetric information and learning: evidence from the automobile insurance market. The Review of Economics and Statistics 87, 197-207.

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

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

Dionne, G., Gourieroux C., Vanasse, C. 2001. Testing for evidence of adverse selection in the automobile insurance market: A comment. The Journal of Political Economy 109, 444-453.

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.

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

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.

Saito, K., 2009. Testing for asymmetric information in the automobile insurance market under rate regulation. The Journal of Risk and Insurance 73, 335-356.

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 the automobile insurance: evidence from Germany. Working Paper.

Su, L., Spindler, M., 2011. Nonparametric testing for asymmetric information. Working Paper.

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

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`.

Chapter 5

Asymmetric Information and Unobserved Heterogeneity: The Case of Accident Insurance

5.1 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 to test theoretical predictions and to depict new directions for research.[1] Two shortcomings are currently still present in this emerging field: One is that many insurance branches have not been analyzed yet, although recent studies show that there is no general answer if there is asymmetric information in insurance markets or not. It depends on the insurance resp. risk under consideration and 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 superimpose the selection on the risk type and this might lead to a negative or zero correlation despite adverse selection.

We contribute to the literature in both directions: Firstly, we analyze the accident insurance which - as far as we know - has never been analyzed before in the literature, although this kind of insurance covers risks which are really essential. Secondly, we try to control for heterogeneity, e.g., differences in risk aversion, by applying finite mixture models.

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

The paper is structured as follows: In section 2 we give a short introduction to the theory of asymmetric information and the principles of the testing procedures. In section 3 the accident insurance is introduced. Before we present the results (section 6) the data set (section 4) and the applied testing procedures (section 5) are explained. Finally, we conclude in section 7.

5.2 The Theory of Asymmetric Information and the Basic Testing Procedures

Asymmetric information comprises two different phenomena, adverse selection and moral hazard.[2] Many equilibrium models of asymmetric information predict that insurees with more insurance coverage should be more likely to experience a loss, i.e., 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 cautious behavior and therefore the expected loss is increased after signing of an insurance contract.
Adverse selection means that the insured knows his risk type ex ante, i.e., before the 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 like 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 if there is positive correlation one has to set up two equations, one for the coverage (C_i) and one for the risk resp. 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 models:

$$C_i = X_i * \beta + \varepsilon_i \tag{2.1}$$

$$L_i = X_i * \gamma + \eta_i \tag{2.2}$$

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 for asymmetric information.

[2] Winter(2000) and Dionne et al. (2000) give surveys over both phenomena.

This kind of test for positive correlation has been widely used in the literature, Cutler and Zeckhauser (2000) review studies in health economics, Cohen and Siegelman (2010) give an overview over results for many different insurance branches. 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 or even negative correlation in insurance markets. Coined 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 risk type (T_i) and about risk aversion (A_i) both influence the unobservable error terms from above:[3]

$$\varepsilon_i = T_i * \kappa_1 + A_i * \kappa_2 + \nu_i \tag{2.3}$$

$$\eta_i = T_i * \xi_1 + A_i * \xi_2 + \tau_i, \tag{2.4}$$

where ν_i and τ_i denote error terms. The principle of the correlation test is that in 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. We will explain these testing procedures in the following sections.

5.3 The Accident Insurance

The accident insurance is very elementary for many insurees as it covers risks which touch the existence of individuals. In Germany there are two pillars concerning the accident insurance:

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

a compulsory accident insurance and a voluntary private accident insurance. The statutory accident insurance covers only risks which are related to the workplace. These are risks like working and commuting accidents and occupational diseases. In the case of such an event this insurance covers the costs for reconstituting the health or pays a pension in the case of incapacity for work. As the compulsory accident insurance applies only to employees (e.g., not to self-employed), as the indemnity sums from the compulsory accident insurance are very limited in the case of an accident and - as mentioned before - apply only to work accidents, a private accident insurance is a very important supplement.

The (private) accident insurance usually covers the following eventualities: invalidity and dismemberment, death, and a hospital per diem[4] in the case of a hospital stay. These are the basic risks which are covered by default. Besides, some insurance companies offer additional insurance payments, e.g., reimbursements of costs for treatment at a health resort or costs for plastic surgery. It is important to mention that the eventualities are only covered if they arise as a consequence of an accident.

For each eventuality the insuree can choose the level of payments in case of occurrence of the event insured. The premium depends on these chosen levels. Additionally, the insurance companies can use variables like occupation, sex, age and so on for risk classification and rating. These observables also determine the insurance premium. In the case of invalidity the insuree chooses the level (i.e. insured sum) 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 invalidity case 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 accident insurance offers two additional characteristics which distinguishes its analysis from, e.g., the automobile insurance:

Disentangling 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 section data the positive correlation cannot be broken down in its constituent parts. It seems reasonable to assume that an accident insurance does not lower the diligence and prudence of the insureds as the consequences are in any case dramatic. For example

[4]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.

in health economics it is often assumed that ex ante moral hazard is negligible. Additionally, cases in which the insured event is caused by purpose or gross negligence are extremely seldom.[5] Therefore if one wants to maintain this assumption the accident insurance offers the opportunity to test only for adverse selection. Otherwise we still offer an analysis of the aggregate effects of asymmetric information.

Accidents vs. claims In the case of insurance contracts with deductibles like in the automobile insurance the distinction between accidents and claims is important. Accidents which are below the deductible are usually not filed and this kind of accidents are 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. As in the casualty insurance all accidents are filed the "accidents vs. claim problem"does not exist here in contrast to many other classes of insurance which have been analyzed recently in the literature.

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

5.4 The Data Set

For our analysis we have 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 claim history up to termination are recorded.

Under one contract several different persons can be insured. For example a father can take out a policy for him, his wife and his kids. Such constellations are also captured in the database.

[5] Personal communication with actuaries.

For each insured risk personal data and claim history are filed. As we want to test the risk assessment of the individual for itself, not the ability to assess the risk of other individuals, we restrict to contracts under which only one person is insured.
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 appr. 2.5 million contract years. This corresponds to $n = 957,506$ insurees.
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 set "historic"and is not in the sample. Therefore there is a sample selection with a tendency for "good"risks being in the sample given the same date of signing ("attrition bias").
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 has still a size of $n = 77,125$. It is reasonable that the individuals might have an informational advantage concerning their accident probability but not concerning the timing of an accident.[6] Therefore this procedure should not influence the results or distort them.

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

Table 5.1: 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
insurance sum for rent	0	165	1075
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 $n = 957,506$

Table 5.2: Number of claims

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

Table 5.3: Descriptive statistics of the subsample

Variable	Minimum	Mean	Maximum
age	14.6	40.3	97
insurance sum for disability	3,835	103,823	1,032,500
insurance sum for death	767	5,415	51,130
insurance sum for hospital per diem	2.5	8.3	65
cost of claims for disability	0	13.9	115,000
cost of claims for death	0	1.29	21000
cost of claims for hospital per diem	0	1.9	5,458

sample size $n = 77,125$

5.5 Methods

In this section we give a short review of the applied methods and customize them to the accident insurance. First, we present the tests for positive correlation introduced by Chiappori and Salanié (2000). This will be the reference point. Second, we give a summary of the test of unused observables (Finkelstein and Poterba (2006)) and finally we introduce finite mixture models.

In the following X denotes the exogenous variables which are used for risk classification by the insurance company. Y denotes the chosen contract. In the accident 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 % invalidity (together with the chosen progression resp. factor) or the daily payment for a hospital stay. All other conditions of the contracts are identical. Z measures the risk. The risk is measured as “ex post risk”, e.g., by the number of accidents or the caused damage payments by the insuree. An index i refers to a certain individual resp. contract which is omitted if there is no confusion.

5.5.1 Testing for a Positive Correlation

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. 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.

5.5.1.1 Two Probits

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

$$\begin{cases} Y_i = \mathbf{1}(X_i\beta + \varepsilon_i > 0) \\ Z_i = \mathbf{1}(X_i\gamma + \eta_i > 0) \end{cases} \tag{5.1}$$

where ε_i and η_i are independent standard normal errors, and β and γ are coefficient vectors (as columns). Y_i is defined as 1, if the insured sum is above a cut off value and 0 if equal or below. Z_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 $\hat{\eta}_i$[7] are calculated. These are required for the following test statistic

$$W_n = \frac{\left(\sum_{i=1}^{n} \hat{\varepsilon}_i \hat{\eta}_i\right)^2}{\sum_{i=1}^{n} \hat{\varepsilon}_i^2 \hat{\eta}_i^2}. \tag{5.2}$$

Under the null of conditional independence, $\text{cov}(\varepsilon_i, \eta_i) = 0$ 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 seldom. But for coverage, Y_i one must choose a cut off point to define the low and high insured sums. In order to check for robustness of our results we will choose several cut off points to see how the definition of the binary variable Y_i influences the results.

5.5.1.2 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- oder LR-test can be

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

used.

5.5.2 Test with Unused Observables

The test for asymmetric information was introduced by Finkelstein and Poterba (2006). The basic principle of this test is that the existence of a variable resp. characteristic that is known to 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 for 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 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:[8]

$$C_i = X_i * \beta + W_i * \alpha + \varepsilon_i \tag{5.3}$$

$$L_i = X_i * \gamma + W_i * \delta + \eta_i \tag{5.4}$$

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 coverage, risk and the exogenous variables used for pricing by the insurance company, and additionally we need variables which are contained in the data but not used for risk classification / pricing (so called unused observables). When there are one or several unused variables and for each we cannot reject the null hypothesis of joint insignificance than 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.
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 for adverse selection and moral hazard as the unique source for the asymmetric information can be excluded.

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

5.5.3 Finite Mixture Models

In this section we briefly introduce 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 we refer to McLachlan and Peel (2000) and 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

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

$\psi = (\alpha, \theta_1, \ldots, \theta_K)$ 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 the concomitant variables and thus constant.

The most common method for maximum likelihood (ml) estimation of finite mixture models with known number of components is the EM algorithm. The EM algorithm was originally introduced for ml estimation of incomplete data (Dempster 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} (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 maximising

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

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. Therefore many variants of the EM algorithm have been proposed in the literature to overcome these shortcomings. Two popular versions, which are also implemented in flexmix, are the 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 we refer to the literature.

For our data set at hand we consider a mixture of bivariate probit models. For the bivariate probits we use the parametrization proposed in Greene (2008).

5.6 Results

We present only the results for the eventuality invalidity. The danger of invalidity is the main reason why insurees buy accident insurance. In the case of the accident insurance only sex and the riskiness of the occupation which is given by two categories are used for pricing. Other variables like age, place of residence and so on are not used although they are available. Therefore we conduct the standard test procedures both only on 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 takes additionally family status, occupation, age and if the insuree has chosen the option to increase the insured sum automatically on an annual basis (annual adjustment). These additional variables allow us also to apply the "unused observables test".

5.6.1 Basic Parametric Tests

In order to apply the introduced parametric procedures one has to define binary variables for "risk" and "coverage".[9] Risk is measured as ex post risk and is defined as 1 if there was at least one accident resp. claim and 0 otherwise, i.e., if there has been no accident during this period. 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 if below or equal to this threshold. In order to check for robustness and to analyze if resp. how the asymmetric information varies with the level of the insured sum we variegate this cut off level. The contracts are identical, only the levels of insured sums can be chosen by the insuree. In Tables 5.4 and 5.5 below the results for the basic and extended setting are given according to different cut off values for the proxy of the insured sum.

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.[10] For the low cut off values 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 already 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 apply tests relying on two probits, on a bivariate probit and a nonparametric test introduced in Chiappori and Salanié (2000) for the basic setting. As all variables used for risk classification are binary in the basic configuration this test is a natural choice. All these tests give consistent results as can be seen in the table.

One important finding is that additional information might reduce resp. eliminate the observed asymmetric information in the accident insurance.

[9] 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 restriction to the correlation test is no limitation.

[10] In the extended version we included age also as year dummies but this has only a slight influence on the results.

Table 5.4: Results - basic configuration

	cut off point in Euro	10,000	50,000	100,000	200,000	300,000	400,000
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
simple 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.

Table 5.5: Results - extended configuration (sex, occupation, family, age, dynamic)

	cut off point in Euro	10,000	50,000	100,000	200,000	300,000	400,000
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.

5.6.2 Test with Unused Observables

In this section we present the results of the test with unused observables. 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 like age, family status, annual adjustment of the insured sum and so on. Therefore we are in the rare situation to have additional information and to conduct the test with unused observables.

There are several possibilities to implement this test. One possibility is to set up two probits, one for the risk and one for the choice of coverage. As regressors we choose in both equations the variables which are used by the insurance company for risk classification and additionally we add one of the unused characteristics. The dependent variables for risk and choice of insurance sum are defined as in the section before. In order to define the binary proxy for the coverage we have to choose a certain cut off value. To check for robustness we choose different cut off values.

An alternative approach is to model the equation for risk by a Cox proportional-hazards regression. In our data set the time of the occurrence of accidents is filed and therefore this can also be interpreted as survival data. Introductions for 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 as

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

or equivalently,

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

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. The choice of coverage is modeled as a probit.

In Table 5.6 below we present the results for both specifications, i.e., modeling the risk equation as probit and as survival model, and for different cut off values in the coverage equation. We find two unused observables which have an significant influence on both variables, age and surprisingly, the choice for 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 which are included as regressors, i.e., a constant, sex and risk of occupation class but they are available upon request. We present only the results for the additionally taken up unused observable in the table.

The estimated survival model is independent of the choice equation, but we repeat the results for each cut off value to see the results for both equations one below the other. The variable annual adjustment (“dynamic”) was coded in the following way: 0 if this option was chosen and 1 otherwise. Therefore the parameters for dynamic in the survival model are interpreted in the following way: Thus for example, holding the other covariates constant, not choosing the annual adjustment reduces the hazard of having an accident by a factor of $e^{-0.543} \approx 0.60$ on average - that is by approximately 40 %.

Table 5.6: Results - unused observables

Unused Variable	Cut off	Equ.	Parameter	t-value
bivariate probit				
age	50,000	risk	−0.002	−1.43
		cov.	−0.020	−68.6***
	100,000	risk	−0.002	−1.79*
		cov.	−0.021	−69.6***
	200,000	risk	−0.002	−1.85*
		cov.	−0.012	−37.5***
"dynamic"	50,000	risk	−0.428	−12.5***
		cov.	−3.038	−123.7***
	100,000	risk	−0.443	−13.0***
		cov.	−2.146	−167.5***
	200,000	risk	−0.426	−12.5***
		cov.	−1.281	−114.5***
survival model				
age	50,000	risk	−0.014	−4.73***
		cov.	−0.020	−68.6***
	100,000	risk	−0.014	−4.73***
		cov.	−0.21	−69.6***
	200,000	risk	−0.014	−4.73***
		cov.	−0.012	−37.4***
"dynamic"	50,000	risk	−0.543	−5.47***
		cov.	−3.041	−123.5***
	100,000	risk	−0.543	−5.47***
		cov.	−2.148	−167.5***
	200,000	risk	−0.542	−5.47***
		cov.	−1.281	−114.5***

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

One could suppose that the age has a significant influence on the choice of coverage and also on the risk. For risk this is the case when we use the survival model, for the probit the characteristic age is only at the edge to become significant. For the choice of coverage age has a significant influence.

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.

5.6.3 Results of the Finite Mixture Models

In this section we present the results of the finite mixture models. For each component we used a bivariate probit model. We implemented the mixture model in R using the package flexmix.[11] As the EM algorithm is only capable of finding local maxima we used different starting values (random assignment, $10 - 20$ trials). We repeated the procedure for different cut off values to check for robustness and to compare the results with the results of a single component.

Each component consists of a bivariate probit (risk and coverage) and we present the corresponding parameter estimates stacked in a column in Table 5.7. 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. In the last row we present the AIC and for all three models it is lower than the corresponding value for the one component bivariate probit model. Therefore the two component model is preferable to a one component model from the point of model selection.

For the models with thresholds $50,000$ and $100,000$ Euro we observe quite similar patterns resp. components: We can identify a group which has a lower risk to have an accident (Comp. 1), while Comp. 2 has a higher risk to have 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 (Comp. 1) than the other. The choice of contract is similar for both groups. The only difference is that for Comp. 2 we have a zero correlation, while for Comp. 1 the correlation is significantly positive.

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

Table 5.7: Results - two components

	cut off 50,000		cut off 100,000		cut off 200,000	
	Comp. 1	Comp. 2	Comp. 1	Comp. 2	Comp. 1	Comp. 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.

5.7 Conclusion

In this paper we analyzed the extent of asymmetric information in the accident insurance especially taking into account unobserved heterogeneity. While the accident insurance covers various eventualities we concentrate on invalidity as it is the main reason for choosing this kind of insurance and covers risks which are essential for the insured. Beside the standard tests for asymmetric information we also apply the test with unused observables and use a finite mixture model with two components. As there are used relatively few variables for risk classification in the accident insurance we are in the rare situation to find variables in the data set which are not used by the insurance companies. We find that the unused observables age and surprisingly the option to choose 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). 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 for the definition of 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 find a significant positive correlation of risk and coverage and therefore we can confirm the result of the standard model especially for low and middle range insured sums. For the high insured sum the correlation within each group is lower and for one component we find a zero correlation between risk and coverage, therefore 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.

In future research we would like to analyze more deeply the joint distribution and dependence structure of risk and coverage and to refine the finite mixture approach, e.g., by applying it to the extended set of variables and to check for stability.

5.8 References

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

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

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, 185-244. Kluwer Academic Publishers.

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 Series in Statistics.

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.

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/

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 Series in Probability and Statistics.

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

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 Series in Probability and Statistics.

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.

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. Working Paper.

Su, L., Spindler, M., 2011. Nonparametric testing for asymmetric information. Working Paper.

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

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`.