Insurance: Mathematics and Economics 49 (2011) 226–239

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

A copula approach to test asymmetric information with applications to predictive modeling

Peng Shi [a,*], Emiliano A. Valdez [b]

[a] Division of Statistics, Northern Illinois University, DuSable Hall 366, DeKalb, IL 60115, USA
[b] Department of Mathematics, University of Connecticut, Storrs, CT 06269-3009, USA

ARTICLE INFO

Article history:
Received September 2010
Received in revised form
April 2011
Accepted 8 April 2011

JEL classification:
C35
D82

Subject Category:
IE30
IM01
IM31

Keywords:
Adverse selection
Moral hazard
Insurance
Copula models
Predictive models

ABSTRACT

In this article, we present a copula regression model for testing asymmetric information as well as for predictive modeling applications in automobile insurance market. We use the Frank copula to jointly model the type of coverage and the number of accidents, with the dependence parameter providing for evidence of the relationship between the choice of coverage and the frequency of accidents. This dependence therefore provides an indication of the presence (or absence) of asymmetric information. The type of coverage is in some sense ordered so that coverage with higher ordinals indicate the most comprehensive coverage. Henceforth, a positive relationship would indicate that more coverage is chosen by high risk policyholders, and vice versa. This presence of asymmetric information could be due to either adverse selection or moral hazard, a distinction often made in the economics or insurance literature, or both. We calibrated our copula model using a one-year cross-sectional observation of claims arising from a major automobile insurer in Singapore. Our estimation results indicate a significant positive coverage–risk relationship. However, when we correct for the bias resulting from possible underreporting of accidents, we find that the positive association vanishes. We further used our estimated model for other possible actuarial applications. In particular, we are able to demonstrate the effect of coverage choice on the incidence of accidents, and based on which, the pure premium is derived. In general, a positive margin is observed when compared with the gross premium available in our empirical database.

1. Background and motivation

In contract theory, asymmetric information is defined to be the situation when one party of an economic transaction possesses information that is not available to the other contractual party. A clear example can be found in the market for used items or goods such as those found on eBay where there is transfer of ownership from one person to another. For such transactions, there are variables, usually observable only during the period of ownership of the goods and oftentimes difficult to trace, to help assist the buyer in assessing the quality of the goods being purchased. As a consequence, the original owner knows the history of the performance of the goods during its period of wear and tear which may be unavailable to the buyer. A classic example can also be found in the market for "used cars". In the seminal work of Akerlof (1970), the economist George A. Akerlof explained the problems arising from information asymmetry based on the "used car" market and referred to defective used cars as "lemons".

The presence of asymmetric information is not strange in the insurance market, and this is true for several forms of insurance (e.g. life, medical, dental, homeowners, and automobile). See Cohen and Siegelman (2010) for a recent review. There is additional information about the probability of a loss unavailable to the insurance company, preventing it from creating a portfolio of homogeneous risks and thus making it more difficult for the insurer to effectively price-discriminate buyers of insurance policies. The economics literature distinguishes asymmetric information due to adverse selection and due to moral hazard. Within the context of insurance, there is presence of *adverse selection* when an insurer does not have enough information to assess those buyers who are considered 'high risks' and who are more likely to purchase insurance. On the other hand, there is presence of *moral hazard* when the behavior of policyholders is altered because they have insurance coverage. Prominent examples of this are in medical insurance when policyholders tend to be less careful of their health and in automobile insurance when policyholders tend to behave recklessly while driving on the road. In this article, because our data

* Corresponding author. Tel.: +1 815 753 6795; fax: +1 815 753 6776.
E-mail addresses: pshi@niu.edu (P. Shi), emiliano.valdez@uconn.edu (E.A. Valdez).

0167-6687/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.insmatheco.2011.04.002

makes it difficult to make the separation, our model for investigating the presence of asymmetric information does not allow for such distinction.

The implications of the presence of asymmetric information can have alarming effects on the insurance market such as possible deterioration of premium adequacy which may lead to insurer insolvency and subsequent market collapse. In combating for asymmetric information, insurers have tended to modify insurance policy designs within the limits of the law. Some of these popular modifications are the introduction of policy deductibles, coinsurance, and policy limits; all tend to have the characteristics that result in the sharing of the risks between the insured and the insurer. In the pioneering work of Rothschild and Stiglitz (1976), the authors demonstrated that when an insurance company offers a basket of products with varying levels of coverage, a separating equilibrium results, in which optimally, as a result, the higher risk chooses a higher level of coverage and pays a higher premium while the lower risk chooses a lower level of coverage and pays a lower premium. As discussed earlier, this phenomenon is called adverse selection. A possible implication of the work of Rothschild and Stiglitz (1976) is that the existence of adverse selection could be tested empirically by examining whether there is a positive relationship between the risk of policyholders and their subsequent selection of the level of insurance coverage.

Since the earliest contribution by Rothschild and Stiglitz (1976), theoretical models for asymmetric information have been extensively developed and extended. We refer readers to the excellent review article by Dionne et al. (2000) for more details. However, a majority of the work on the empirical testing of asymmetric information in insurance markets has been centered on the coverage–risk relationship (see Chiappori et al. (2006) and Cohen and Siegelman (2010)). Hoping to contribute to the literature along this line, we focus on the automobile insurance market primarily because we are motivated by the available data we have for investigation and by the fact that there have been several empirical investigations in this area making it easy to compare our results with existing ones. Although theory generally suggests the presence of asymmetric information, earlier empirical investigations, however, have rather been inconclusive; some studies found strong evidence while others have suggested no evidence at all. Several of these studies share many model features such as the use of a policyholder's claim history to measure the risk level of the driver in addition to the effects of the choice of policy coverage (the presence of deductibles, coinsurance and/or policy limits). Moreover, several of the empirical models control for certain observable characteristics that allow the insurer to classify the risk level of the policyholders. However, several of these studies are distinct in many aspects. First is the choice of the risk level of the driver: some treat the likelihood of the event that an accident, or subsequently a claim, occurs while others use the number of times an insured has made a claim as the primary variable of interest. Next of course are the differences in the empirical data used in the investigation itself where other controlling effects make the comparison non-uniform. Finally and obviously are the differences in the model choice, which range from using logit or probit models for the probability of a claim to using various count data regression models for the number of accidents or claims.

To illustrate, using an ordered logit model, Puelz and Snow (1994) found a strong positive correlation between the number of times an insured has made a claim together with the level of coverage purchased accounting for the risk class to which the insured has been classified. Based on the notion of conditional dependence, Dionne et al. (2001) suggested that the presence of adverse selection vanishes when "the nonlinearity of the risk classification variables" are taken into account. Using a portfolio of contracts from a French insurer, Chiappori and Salanié (2000) also concluded no evidence of asymmetric information in the automobile insurance market. The authors proposed the use of a bivariate probit regression to jointly model the choice of a better coverage and the occurrence of an accident. The rationale is that the influence between the occurrence of accidents and the policy selection could go in either direction. Such method has also been employed by Cohen (2005) using data from an insurer that operates in Israel, and by Saito (2006) using data from a Japanese automobile insurer, to test the asymmetric information in the automobile insurance market. These articles surprisingly have opposing results; Cohen (2005) concluded there is evidence of information asymmetry and Saito (2006) found no adverse selection or moral hazard in this market.

As earlier stated, the adverse selection effect implies that a higher risk, one considered to have a higher chance of accident occurrence, chooses a more comprehensive and better coverage. However, the occurrence of accident could be the result of a moral hazard effect, that is, the selection of a more comprehensive and better coverage may provide a motivation to be a careless driver leading back to a higher chance of getting into an accident. Consequently, in the literature, the coverage–occurrence relation has been attributed to a combined effect of adverse selection and moral hazard. Kim et al. (2009) recently pointed out that dichotomous indicators might not completely reveal the policyholder's selection decision. Henceforth, a multinomial measure was suggested for the policyholder's coverage selection. Similar to Richaudeau (1999), these authors considered a count data regression model to examine the effect of policy selection on the occurrence of accidents.

Another factor that may affect the coverage–occurrence relationship is the degree of risk aversion of policyholders. An individual with a higher aversion to risk is more likely to purchase better coverage and even be considered to be a more cautious driver. In contrast with adverse selection and moral hazard, this observation implies a negative relationship between occurrence of accident and choice of coverage. Such negative correlation could offset the positive correlation implied by the traditional adverse selection. Research work along this line include deMeza and Webb (2001) and Chiappori et al. (2006), where the authors argue that the absence of the positive correlation between the level of risk and the choice of coverage could be consistent with the presence of asymmetric information. Thus, the empirical results in our study represent a combined effect of various types of asymmetric information.

We note a common limitation shared by our method and other published empirical studies based on cross-sectional observations, that the coverage–occurrence relationship represents a combined effect of various types of asymmetric information. As pointed out by Abbring et al. (2003b), a longitudinal observation is useful in distinguishing between adverse selection and moral hazard. Some recent studies in this line include Abbring et al. (2003a), Dionne et al. (2005), Israel (2007), and Dionne et al. (2010). Another related group of studies on dynamic testing involves learning, where either policyholders or insurers, or even both, become able to understand the degrees of risks over time, for example, see Israel (2006) and Cohen (2005, 2008).

Motivated by examining the asymmetric information in the automobile insurance market, we propose to use a copula function to jointly model the policyholder's choice of coverage and occurrence of accident. Our approach differs from existing studies in several respects. Unlike linear correlation models used in previous studies, the copula approach flexibly allows to capture both linear and nonlinear coverage–risk relationships. By modeling the two responses simultaneously, our method avoids the potential endogeneity issue that possibly arises when examining the effect of a multinomial coverage selection measure on the number of accidents. Although we find this method to be