Nonparametric Testing for Asymmetric Information

Liangjun Su[a], Martin Spindler[b]
[a] School of Economics, Singapore Management University (ljsu@smu.edu.sg)
[b] Ludwig–Maximilians–University and Max Planck Society, Germany
(spindler@mea.mpisoc.mpg.de)

First version: July 23, 2010
This version: November 28, 2012

Abstract

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 in the last two decades. Almost all testing procedures that are used in empirical studies are parametric, which may yield misleading conclusions in the case of misspecification of either *functional* or *distributional* relationships among the variables of interest. Motivated by the literature on testing conditional independence, we propose a new nonparametric test for asymmetric information which is applicable in a variety of situations. We demonstrate that the test works reasonably well through Monte Carlo simulations and apply it to an automobile insurance data set and a long term care insurance data set. Our empirical results consolidate Chiappori and Salanié's (2000) findings that there is no evidence for the presence of asymmetric information in the French automobile insurance market. While Finkelstein and McGarry (2006) find no positive correlation between risk and coverage in the long term care insurance market in the US, our test detects asymmetric information using only the information that is available to the insurance company and our investigation of the source of asymmetric information suggests some sort of asymmetric information that is related to risk preferences as opposed to risk types and thus lends support to Finkelstein and McGarry (2006).

Keywords: Asymmetric information, Automobile insurance, Long term care insurance, Conditional independence, Distributional misspecification, Functional misspecification, Nonlinearity, Nonparametric test.

JEL classification codes: C12, C14, D82, D86, G22.

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 be described exhaustively 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 of 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 hazard 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 literature. Nevertheless, these parametric tests are fragile to both *functional* and *distributional* form misspecifications. In actuarial science the number of claims is modeled in many cases by a Poisson distribution or a negative binomial distribution. The size of claims is modeled by a generalized gamma distribution, a log normal distribution, or even more elaborate distributions. For an overview of modeling of claims in actuarial science, see Kaas et al. (2008), Mikosch (2008) and De Jong and Heller (2008), among others. They show that modeling claims by a probit model might be insufficient and more sophisticated models and methods are used by actuaries. 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 on testing for asymmetric information. 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). Simulations reveal that our test behaves well in finite samples in comparison with some existing tests. We then apply our test to a French automobile insurance data set and a US long term insurance data set and compare our testing results with the results found in the literature. Our empirical results consolidate Chiappori and Salanié's (2000) findings that there is no evidence for the presence of asymmetric information in the French automobile insurance market. While Finkelstein and McGarry (2006) find no positive correlation between risk and coverage in the long term care insurance market, our test reveals evidence of asymmetric information using only the information that is available to the insurance company.

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. In Section 6 we test for asymmetric information in the two applications mentioned above. Final remarks are contained in Section 7. All technical details are relegated to the Appendix.

2 The Theory of Asymmetric Information

In their seminal paper Rothschild and Stiglitz (1976) introduce the notion of adverse selection in insurance markets, which has since then been extended in many directions. For a detailed survey on adverse selection and the related moral hazard problem, we refer to Dionne, Doherty and Fombaron (2000) and Winter (2000), respectively. 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. 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 levels 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, Salanié, and Salanié (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. To disentangle moral hazard from adverse selection is undoubtedly an important problem in the empirical literature. See Dionne, Michaud, and Dahchour (2012) for the first attempt and Cohen and Siegelman (2010) for an overview of different possible strategies for dealing with this problem.

In sum, the theory of asymmetric information 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 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. For the theory of risk classification under asymmetric information we refer to the survey article Crocker and Snow (2000). Only within each class are the mechanisms described above valid.

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, which has been first proposed by Dionne, Gourieroux, and Vanasse (2001, 2006), and then review the parametric and nonparametric testing procedures in turn.

3.1 General Structure

In the following we denote by X the vector of exogenous control variables to be conditioned 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 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. 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. 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. As Dionne, Gourieroux, and Vanasse (2001, 2006) first note, this can be formally 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 a contract, e.g., the choice of certain deductible, provides no useful information for predicting the risk, e.g., the number of claims, as long 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.

As a referee kindly points out, there are applications in different insurance markets that use more general models where all variables can be continuous as in Dionne, Gourieroux, and Vanasse (2001, 2006). In this case, one can apply some existent tests for conditional independence (e.g., Delgado and González-Manteiga (2001), Su and White (2007, 2008)) to test for asymmetric information.

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 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 (β_1, β_2') are intercept and slope coefficients, respectively, 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, Gourieroux, and Vanasse (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 $\hat{\eta}_i$, which have been introduced by Gourieroux et al. (1987), and then construct the following test statistic

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

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

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

$$\mathbb{H}_0 : F(Z|X, Y) = F(Z|X) \text{ a.s.}, \tag{4.1}$$

and the alternative hypothesis is

$$\mathbb{H}_1 : \Pr\{F(Z|X, Y) = F(Z|X)\} < 1. \tag{4.2}$$

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 focuses on the case where both Y and X are continuously distributed, see, Delgado and González-Manteiga (2001), Su and White (2007, 2008), Song (2009), Huang and White (2009), Huang (2010), 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.

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 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 $F_{Z|X}(z|x)$ (resp. $F_{Z|X,Y}(z|x, y)$) to denote the conditional CDF of Z given X (resp. (X, Y)). Below we make reference to these CDFs and several probability density functions (PDFs) by simply using the list of their arguments – for example, $f(x, y, z)$, $f(x, y)$ and $f(x)$ 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 $X_i = (X_i^{c\prime}, X_i^{d\prime})'$ 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. When some of the conditioning variables in X_i have a natural ordering, one can easily modify the corresponding discrete kernel defined below following either Racine and Li (2004) or Li and Racine (2007, 2008). We use X_{is}^c (resp. X_{is}^d) to denote the sth component of X_i^c (resp. X_i^d), where $s = 1, \cdots, p_c$ (resp. 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 constant (Nadaraya-Watson) method. [Alternatively, one can consider local linear/polynomial method, but we find through simulations that the latter method does not yield as satisfactory size behavior as the local constant 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, $x^c \equiv (x_1^c, \cdots, x_{p_c}^c)'$. 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 the vector of discrete variables 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:

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

We estimate $F\left(Z_i|X_i, y\right)$ by

$$\widehat{F}\left(Z_i|X_i, y\right) = \frac{\frac{1}{n}\sum_{j=1}^n K_{h\lambda,ji} \mathbf{1}_j^y \mathbf{1}\left(Z_j \leq Z_i\right)}{\frac{1}{n}\sum_{j=1}^n K_{h\lambda,ji} \mathbf{1}_j^y} \tag{4.7}$$

where $\mathbf{1}_j^y \equiv \mathbf{1}\left(Y_j = y\right)$. Then we measure the variation in $\widehat{F}\left(Z_i|X_i, y\right)$ across different values of y and different observations by

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

where $a(\cdot)$ is a uniformly bounded nonnegative weight function with compact support $\mathcal{X}^c$ that lies in the interior of the support of X_i^c. It serves to perform trimming in areas of sparse support. In the following simulations and empirical studies, we will set $a\left(X_i^c\right) = \Pi_{j=1}^{p_c} \mathbf{1}(q_j(0.025) \leq X_{ij}^c \leq q_j(0.975))$ where $q_j(\tau)$ denotes the τth sample quantile of X_{ij}^c. We will study the asymptotic properties of D_n under $\mathbb{H}_0$, a sequence of Pitman local alternatives, and the global alternative $\mathbb{H}_1$. 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.

4.2 Assumptions

Throughout the paper we use ξ_i and ς_i to denote $(X_i', Y_i, Z_i)'$ and $(X_i', Z_i)'$, respectively. Similarly, let $\xi \equiv (x', y, z)'$ and $\varsigma \equiv (x', z)'$. With a little bit abuse of notation, we use $f(\xi)$, $f(x, y)$, and $f(x)$ to denote the PDF of ξ_i, (X_i, Y_i), and X_i, respectively. Frequently we also write $F(z|x, y)$ as $F(z|x^c, x^d, y)$.

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

Assumption A1. The sequence $\{\xi_i\}_{i=1}^n$ is independent and identically distributed (IID) with CDF F_ξ.

Assumption A2. $f(\xi)$ is uniformly bounded on $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, 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. $f(x, y)$ is bounded away from 0 on $\mathcal{X} \times \mathcal{Y}$.

Assumption A3. (i) For each $(x^d, y) \in \mathcal{X}^d \times \mathcal{Y}$ and $z \in \mathcal{Z}$, $F(z|x^c, x^d, y)$ has all partial derivatives up to order 2 with respect to x^c, and its second order partial derivatives with respect to x^c, $\partial^2 F(z|x^c, x^d, y)/\partial x_s^c \partial x_t^c$, $s, t = 1, \cdots, p_c$, are uniformly continuous and bounded on $\mathcal{X}^c$.

(ii) For each $(x, y) \in \mathcal{X} \times \mathcal{Y}$, $|F(z|x, y) - F(\widetilde{z}|x, y)| \leq C\ |z - \widetilde{z}|$ for all $z, \widetilde{z} \in \mathcal{Z}$.

Assumption A4. (i) The kernel function $q : \mathbb{R} \to \mathbb{R}^+$ is a continuous, bounded, and symmetric PDF.

(ii) 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 $|q(u) - q(\widetilde{u})| \leq c_2 |u - \widetilde{u}|$ for any $u, \widetilde{u} \in \mathbb{R}$; or $q(\cdot)$ is differentiable, $|dq(u)/du| \leq c_1$, and for some $\iota_0 > 1$, $|dq(u)/du| \leq c_1 |u|^{-\iota_0}$ for all $|u| > c_2$.

Assumption A5. Let $h! \equiv \Pi_{s=1}^{p_c} h_s$. Let $\|\cdot\|$ denotes the Euclidean norm. 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 A1 is standard in cross sectional studies. One could allow dependence but that would complicate the presentation to a large degree. Assumption A2 is standard for kernel estimation with mixed regressors. Assumptions A3-A4 are used to obtain uniform consistency for kernel estimators; see, e.g., Hansen (2008). Assumption A5 imposes appropriate conditions on the bandwidth. In particular A5 implies that undersmoothing is required for our test and $p_c < 4$. This is typical in nonparametric tests when a second order kernel is used in the local constant regression. In the case where $p_c \geq 4$, one has to rely upon the use higher order kernels.

4.3 The Asymptotic Distribution of the Test Statistic

Let $\sigma^2(z|x,r) \equiv \mathrm{Var}(\mathbf{1}(Z_i \leq z) \mid X_i = x, Y_i = r) = F(z|x,r)[1 - F(z|x,r)]$ and $V(z, \overline{z}; x, r) \equiv \mathrm{Cov}(\mathbf{1}(Z_i \leq z), \mathbf{1}(Z_i \leq \overline{z}) \mid X_i = x, Y_i = r) = F(z \wedge \bar{z}|x,r) - F(z|x,r)F(\bar{z}|x,r)$, where $a \wedge b = \min(a,b)$. Define

$$B_0 \equiv \frac{C_1(c_y - 1)}{(h!)^{1/2}} \sum_{r=0}^{c_y-1} \sum_{x^d \in \mathcal{X}^d} \int \int f(x,r)^{-1} \sigma^2(z|x,r) a(x^c) f(x) dF(z|x) dx^c, \tag{4.8}$$

$$\begin{aligned} \sigma_0^2 \equiv\ & 2C_2(c_y - 1)^2 \sum_{r=0}^{c_y-1} \sum_{x^d \in \mathcal{X}^d} \int \int \int f(x,r)^{-2} V(z, \overline{z}; x, r)^2 a(x^c)^2 \\ & \times f(x)^2 dF(z|x) dF(\overline{z}|x) dx^c, \end{aligned} \tag{4.9}$$

where $C_1 = [\int q(u)^2 du]^{p_c}$ and $C_2 = \{\int [\int q(u) q(u-v) du]^2 dv\}^{p_c}$.

Our first result says that after centering, $(h!)^{1/2} D_n$ is asymptotically normally distributed under $\mathbb{H}_0$.

Theorem 4.1 *Suppose Assumptions A.1-A.5 hold. Then under* $\mathbb{H}_0$, $(h!)^{1/2} D_n - B_0 \xrightarrow{d} N(0, \sigma_0^2)$.

To implement the test, we need to consistently estimate B_0 and σ_0^2. Define

$$\widehat{B}_n \equiv \frac{C_1(c_y - 1)}{n(h!)^{1/2}} \sum_{i=1}^{n} \sum_{r=0}^{c_y-1} \widehat{f}(X_i, r)^{-1} \widehat{\sigma}^2(Z_i|X_i, r) a(X_i^c),$$

$$\widehat{\sigma}_n^2 \equiv \frac{2C_2(c_y - 1)^2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{r=0}^{c_y-1} \widehat{\beta}_r(\varsigma_i, \varsigma_j) K_{h\lambda,ji} a(X_i^c) a(X_j^c),$$

where $\widehat{\beta}_r(\varsigma_i, \varsigma_j) = \widehat{f}(X_i, r)^{-1} \widehat{f}(X_j, r)^{-1} \widehat{V}(Z_i, Z_j; X_i, r) \widehat{V}(Z_i, Z_j; X_j, r)$, $\widehat{f}(X_i, r) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}_j^r \times K_{h\lambda,ji}$, $\widehat{\sigma}^2(z|x,r) = \widehat{F}(z|x,r)[1 - \widehat{F}(z|x,r)]$, and $\widehat{V}(z, \overline{z}; x, r) = \widehat{F}(z \wedge \bar{z}|x,r) - \widehat{F}(z|x,r)\widehat{F}(\bar{z}|x,r)$. We demonstrate in the proof of Theorem 4.2 below that $\widehat{B}_n - B_0 = o_P(1)$ and $\widehat{\sigma}_n^2 - \sigma_0^2 = o_P(1)$. Then we can compare

$$T_n \equiv \left[(h!)^{1/2} D_n - \widehat{B}_n\right] / \sqrt{\widehat{\sigma}_n^2} \tag{4.10}$$

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:

$$\mathbb{H}_1(\gamma_n): F(z|x,r) - F(z|x,s) = \gamma_n \delta_{n,rs}(\varsigma) \text{ for a.e. } \varsigma, \tag{4.11}$$

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 4.2 *Suppose Assumptions A1-A5 hold. Then under* $\mathbb{H}_1(\gamma_n)$ *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 $\mathbb{H}_1$ stated in (4.2).

Theorem 4.3 *Suppose Assumptions A1-A5 hold. Then under* $\mathbb{H}_1$, $n^{-1}(h!)^{-1/2}T_n = \mu_A/\sigma_0 + o_P(1)$, *where* $\mu_A \equiv \sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1} E\left[F(Z_i|X_i, r) - F(Z_i|X_i, s)\right]^2$, *so that* $P(T_n > c_n) \to 1$ *under* $\mathbb{H}_1$ *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

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

where $\widetilde{F}(z|x)$ and $\widetilde{F}(z|x, y)$ are local constant 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 (4.7)) and thus avoid the above dilemma.

5 Monte Carlo Simulations

In this section we conduct some Monte Carlo experiments to evaluate the finite sample performance of our test and compare it with some existing tests.

5.1 Data Generating Processes

We consider three data generating processes (DGPs):

DGP 1.

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

where $X_i \equiv \left(X_{i1}^c, X_{i2}^c, X_{i1}^d, X_{i2}^d\right)'$, $\phi(\cdot)$ is the $N(0,1)$ PDF, $d_i = \sqrt{1 + X_{i1}^{c2} + X_{i2}^{c2}}$, X_{i1}^c is IID $N(0,1)$, X_{i2}^c is IID $N(0,1)$, $P\left(X_{i1}^d = l\right) = 1/4$ 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.

$$
\begin{aligned}
Y_i &= \mathbf{1}\left(\varepsilon_{Yi} \leq m_Y\left(X_i\right)\right), \\
Z_i &= m_Z\left(X_i\right) + s\ \varepsilon_{Zi}, \\
m_Z\left(X_i\right) &= \phi\left(X_{i1}^c\right) X_{i2}^c - X_{i1}^c - X_{i2}^c X_{i2}^d + 0.5X_{i1}^d X_{i2}^d + 2\delta Y_i \left(X_{i1}^c\right)^2,
\end{aligned}
$$

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

DGP 3.

$$
\begin{aligned}
Prob\_acc_i &= \Phi\left(-2 + X_i^c + 0.2X_i^d + \delta X_{i,private}^c\right), \\
Premium_i &= \Phi\left((-2 + 0.2X_i^d + X_i^c)(0.6\varepsilon_{Yi} + 0.8)\right), \\
Y_i &= \mathbf{1}\left(Premium_i \leq Prob\_acc_i\right), \\
Z_i &= \mathbf{1}\left(Prob\_acc_i \leq \varepsilon_{Zi}\right),
\end{aligned}
$$

where $\Phi(\cdot)$ is the $N(0,1)$ CDF, X_i^c is IID $N(0,1)$, X_i^d is IID with $P\left(X_i^d = l\right) = 1/2$ for $l = 0$ and 1, $X_{i,private}^c$ is IID $U(-0.5, 0.5)$, ε_{Yi} and ε_{Zi} are each IID $U(0,1)$ and mutually independent.

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 vector X_i includes two continuous variables, X_{i1}^c and X_{i2}^c, and two discrete variables, X_{i1}^d and X_{i2}^d. DGP 3 tries to mimic an insurance market with asymmetric information in a simplistic way. The accident probability of an individual ($Prob\_acc_i$) is determined by three variables, two continuous ones ($X_i^c, X_{i,private}^c$) and a discrete one (X_i^d), indicative of gender, say. While the insuree can observe all three variables, the insurance company can observe only X_i^c and X_i^d. The insurance company tries to assess the individual accident probabilities based on the observed information and requires some mark-up. Without loss of generality, we standardize the cost in the case of an accident to one. An individual buys insurance coverage if it is advantageous for him, i.e., under standardization if the premium is lower than his accident probability. Finally, we simulate whether an accident occurs depending on the individual accident probabilities.

5.2 Tests and Implementation

We consider three tests for conditional independence. The first is the two-probit parametric test (Two-probit) and the second is the nonparametric test proposed by Chiappori and Salanié (2000) (CS's NP test), both of which are described in Section 3.2. We compare the performance of these two tests with our nonparametric test. As the two-probit test and the bivariate probit test tend to yield very similar results, we present only the results of the former test. To implement the two-probit test for DGP 2, we first transform the continuous Z_i variable into a binary one by using the value 0 as a cutoff point.

For DGP 1-2 we consider two different configurations to apply the two-probit test. In Configuration I we consider a "correctly specified "(nonlinear) probit model where all variables in X_i enter the probit equations as specified in DGPs 1-2, but we do not allow the interaction term $Y_i X_{i1}^c$ in DGP 1 or $Y_i \left(X_{i1}^c\right)^2$ in DGP 2 in order to generate residual correlations under the alternative hypothesis (i.e., when $\delta \neq 0$). For either DGP 1 or DGP 2, we consider the following nonlinear probit models:

$$\begin{aligned}
Y_i &= \mathbf{1}\left(\varepsilon_{Yi} \leq \beta_0 + \beta_1 \left(X_{i1}^c/d_i\right) + \beta_2 \left(X_{i1}^{c2}/d_i\right) + \beta_3(\phi\left(X_{i2}^c\right)/d_i) + \beta_4 \left(X_{i1}^c X_{i2}^c/d_i\right)\right. \\
&\quad \left. + \beta_5 \left(X_{i1}^c X_{i1}^d/d_i\right) + \beta_6 \left(X_{i1}^d/d_i\right) + \beta_7 \left(X_{i1}^d X_{i2}^d/d_i\right)\right) \\
Z_i &= \mathbf{1}\left(\varepsilon_{Zi} \leq \gamma_0 + \gamma_1 \phi\left(X_{i1}^c\right) X_{i2}^c + \gamma_2 X_{i1}^c + \gamma_3 X_{i2}^c X_{i2}^d + \gamma_4 X_{i1}^d X_{i2}^d\right)
\end{aligned}$$

where $d_i = \sqrt{1 + X_{i1}^{c2} + X_{i2}^{c2}}$ and ε_{Yi} and ε_{Zi} are each modeled as $N(0,1)$. In Configuration II we consider the widely used linear probit model where all variables in X_i enter the probit equations linearly and the two probit models become

$$\begin{aligned}
Y_i &= \mathbf{1}\left(\varepsilon_{Yi} \leq \beta_0 + \beta_1 X_{i1}^c + \beta_2 X_{i2}^c + \beta_3 X_{i1}^d + \beta_4 X_{i2}^d\right) \\
Z_i &= \mathbf{1}\left(\varepsilon_{Zi} \leq \gamma_0 + \gamma_1 X_{i1}^c + \gamma_2 X_{i2}^c + \gamma_3 X_{i1}^d + \gamma_4 X_{i2}^d\right)
\end{aligned}$$

where ε_{Yi} and ε_{Zi} are each modeled as $N(0,1)$. Note that we have dropped the interaction term associated with Y_i in the probit equation of Z_i. Otherwise there would be symmetric information independently of the true value of δ and no residual correlation can be detected even when δ is nonzero, i.e., the alternative hypothesis holds. The results are reported in columns 4-5 and 6-7 in Table 1 for Configurations I and II, respectively.

For DGP 3 we run two different versions of the two-probit test. In Configuration I we adopt a "full information" approach which also uses the variable $X^c_{i,private}$ in both probit equations:

$$\begin{aligned} Y_i &= \mathbf{1}\left(\varepsilon_{Yi} \leq \beta_0 + \beta_1 X_i^c + \beta_2 X_i^d + \beta_3 X^c_{i,private}\right) \\ Z_i &= \mathbf{1}\left(\varepsilon_{Zi} \leq \gamma_0 + \gamma_1 X_i^c + \gamma_2 X_i^d + \gamma_3 X^c_{i,private}\right). \end{aligned}$$

In Configuration II we use the correct information set of the insurance company and exclude $X^c_{i,private}$ from the above two linear probit equations. The results are reported in columns 4-5 and 6-7 in Table 1 for Configurations I and II, respectively.

To construct our test statistic, we need to choose both kernel and bandwidth. We choose the Gaussian kernel for the continuous regressor(s): $q(x) = (2\pi)^{-1/2}\exp\left(-x^2/2\right)$, and select the bandwidth following the lead of Horowitz and Spokoiny (2001) and Su and Ullah (2009). Specifically, we use a geometric grid consisting of N points $h^{(s)}$, where $h^{(s)} = (h_1^{(s)}, ..., h_l^{(s)})'$, $h_j^{(s)} = \omega^s s_j h_{\min}$, $j = 1, ..., l$ ($l = 2$ in DGPs 1-2 and $=1$ in DGP 3), $s = 0, 1, ..., N-1$, s_j is the sample standard deviations of X^c_{ij} in DGPs 1-2 and X^c_i in DGP 3, respectively, $N = \lfloor \log n \rfloor$, $\lfloor \cdot \rfloor$ denotes the integer part of $\cdot$, and $\omega = (h_{\max}/h_{\min})^{1/(N-1)}$. Note that undersmoothing is required for our tests and the choice of bandwidths depends on the dimension p_c of the continuous variables in X_i. We set $h_{\min} = 0.5n^{-1/4.75}$, $h_{\max} = 3n^{-1/4.75}$, $\lambda_1 = \lambda_2 = 0.5n^{-2/4.75}$ in DGPs 1-2, and $h_{\min} = 0.5n^{-1/4.25}$, $h_{\max} = 5n^{-1/4.25}$, $\lambda_1 = 0.5n^{-2/4.25}$ in DGP 3. Write $\lambda = (\lambda_1, \lambda_2)'$ for DGPs 1-2 and λ_1 for DGP 3. For each $\left(h^{(s)}, \lambda\right)$, we calculate the test statistic in (4.10) and denote it as $\widehat{T}\left(h^{(s)}, \lambda\right)$. Define

$$\mathrm{Sup}T_n \equiv \max_{0 \leq s \leq \mathcal{N}-1} \widehat{T}\left(h^{(s)}, \lambda\right). \tag{5.1}$$

Even though $\widehat{T}\left(h^{(s)}, \lambda\right)$ is asymptotically distributed as $N(0,1)$ under the null for each s, the distribution of $\mathrm{Sup}T_n$ is unknown. Fortunately, we can use 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 $\mathrm{Sup}T_n^*$ in the same way as $\mathrm{Sup}T_n$ by using $\{(X_i^*, Y_i^*, Z_i^*)\}_{i=1}^n$ instead.

4. Repeat steps 1-3 B times to obtain B bootstrap test statistics $\left\{\mathrm{Sup}T_{nj}^*\right\}_{j=1}^B$. Calculate the

bootstrap p-values $p^* \equiv B^{-1} \sum_{j=1}^{B} \mathbf{1}\left(\mathrm{Sup}T_{nj}^* \geq \mathrm{Sup}T_n\right)$ and reject the null hypothesis of conditional independence if $p^*$is smaller than the prescribed level of significance.

The above procedure combines the local bootstrap procedure of Paparoditis and Politis (2000) with the rate-optimal test idea of Horowitz and Spokoiny (2001). 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 $b = s_Z n^{-1/(p_c+4)}$ 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 $\widetilde{h} = (\widetilde{h}_1, ..., \widetilde{h}_{p_c})'$ and $\widetilde{\lambda} = \lambda$ with $\widetilde{h}_j = s_j n^{-1/(4+p_c)}$. Following Horowitz and Spokoiny (2001) one can justify the asymptotic validity of the above bootstrap approximation.

5.3 Test Results

Table 1 reports the empirical rejection frequencies of the three tests at 5% and 10% nominal levels for DGPs 1-3. To save on computational time, we use 250 replications for each sample size n and 200 bootstrap resamples in each replication. We summarize some important findings from Table 1.

First, in terms of level, only our nonparametric smoothing test is well behaved across all DGPs and all sample sizes under investigation. The two-probit test under Configuration I has well behaved level for all DGPs under investigation. The two-probit test under Configuration II tends to be excessively oversized for DGP 1 and moderately oversized for DGP 2. The CS's NP test has the correct size for DGP 3 but severe size distortion for DGPs 1-2. In general when size distortion exists, it tends to be inflated when n increases; see the cases of two-probit test (Configuration II) and CS's NP test for DGPs 1-2.

Second, in terms of power, we find that our test has stable power to detect deviations from conditional independence no matter whether Z_i is discrete or continuous. As either n or δ increases, the power of our nonparametric test increases very quickly. Since the CS's NP test is oversized in DGPs 1-2 and there is no way to obtain a size-corrected power for their test, we cannot compare their test power with ours. When the size of their test is well-behaved in DGP 3, we find that the power of their test is significantly lower than that of ours when the signal of asymmetric information is not very strong ($\delta = 1$). Interestingly, the two-probit test has totally distinct power behavior under the two configurations. Under Configuration I when the correctly specified functional form is used in DGP 1-2 or the full information is explored in DGP 3, there is not much residual information left in the residual so that two-probit test exhibits little power in detecting residual correlation. In sharp contrast, under Configuration II when the functional form in the probit models are not correctly accounted for in DGPs 1-2 or only partial information is explored in DGP 3, the residuals from the two probit models contain much useful information

Table 1: Finite sample rejection frequency for DGPs 1-3

DGP	n	δ	Two-probit (Configuration I)		Two-probit (Configuration II)		CS's NP test		Our test	
			5%	10%	5%	10%	5%	10%	5%	10%
1	200	0	0.056	0.092	0.084	0.164	0.152	0.252	0.044	0.088
		1	0.048	0.104	0.128	0.208	0.508	0.696	0.052	0.168
		2	0.064	0.124	0.120	0.208	0.932	0.968	0.208	0.348
	400	0	0.044	0.108	0.140	0.208	0.276	0.400	0.060	0.108
		1	0.080	0.152	0.212	0.288	0.908	0.956	0.176	0.312
		2	0.100	0.168	0.212	0.372	0.996	1.000	0.664	0.832
	600	0	0.052	0.104	0.224	0.316	0.472	0.612	0.056	0.120
		1	0.072	0.160	0.276	0.412	0.988	0.976	0.412	0.616
		2	0.124	0.192	0.340	0.476	1.000	1.000	0.944	0.976
2	200	0	0.064	0.108	0.052	0.124	0.088	0.160	0.028	0.068
		1	0.040	0.084	0.072	0.144	0.196	0.296	0.084	0.204
		2	0.056	0.116	0.064	0.132	0.352	0.552	0.164	0.280
	400	0	0.068	0.108	0.060	0.120	0.124	0.200	0.036	0.096
		1	0.078	0.072	0.080	0.172	0.412	0.576	0.284	0.440
		2	0.072	0.132	0.108	0.192	0.644	0.784	0.420	0.576
	600	0	0.052	0.112	0.076	0.148	0.176	0.284	0.054	0.112
		1	0.040	0.084	0.080	0.172	0.412	0.576	0.528	0.664
		2	0.056	0.142	0.132	0.204	0.780	0.884	0.672	0.784
3	200	0	0.024	0.096	0.024	0.088	0.052	0.104	0.060	0.112
		1	0.060	0.080	0.380	0.532	0.096	0.156	0.348	0.420
		2	0.028	0.096	0.952	0.980	0.720	0.848	0.896	0.940
	400	0	0.076	0.100	0.076	0.104	0.044	0.108	0.044	0.100
		1	0.068	0.084	0.616	0.712	0.292	0.276	0.508	0.632
		2	0.024	0.088	1.000	1.000	0.984	0.996	1.000	1.000
	600	0	0.056	0.092	0.056	0.092	0.044	0.092	0.064	0.120
		1	0.036	0.096	0.804	0.856	0.412	0.588	0.732	0.796
		2	0.036	0.084	1.000	1.000	1.000	1.000	1.000	1.000

and may exhibit significant residual correlation as expected. Despite the over-size issue of the two-probit test for DGPs 1-2 under Configuration II, its power is not as good as that of our nonparametric test.

6 Empirical Applications

In this section we apply our nonparametric test to an automobile insurance data set and a long term care insurance data set.

6.1 Automobile Insurance

Despite the scarcity of insurance data sets, 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). 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, we believe that our methodology is broadly applicable.

6.1.1 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. 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.

6.1.2 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. 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. In this context "young" refers not to the actual age but to the driving experience. 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 heterogeneity 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. Chiappori and Heckmann (1999) discuss this point in detail. We pursue the second approach and concentrate on novice drivers.

An important issue in testing for asymmetric information is the distinction between accidents and claims. The data set of insurance companies comprise 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. 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.

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 is 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. Table 2 summarizes the key figures of the data

Table 2: Summary statistics for the car insurance data

Sample size	$6,333$
Contracts under TR	$2,335$ (36.9%)
Contracts with one or more accidents	434 (6.9%)

Note: Percentages are in brackets.

set.

For the 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 car (6), type of use (4), type of area (5), profession of driver (8), region (10), age of driver, and age of 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 car or 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 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. 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 car and the age of driver as continuous variables. See Table 3 for a summary of these configurations.

6.1.3 Empirical Results

In Figure 1 we plot the predicted accident probabilities as a function of the two continuous variables, namely, age of car and age of driver, conditional on insurance coverage (i.e., RC or TR contract) by using all variables under Configuration I. We calculate three predictions for the accident probabilities based on a probit model, our nonparametric local constant estimate,

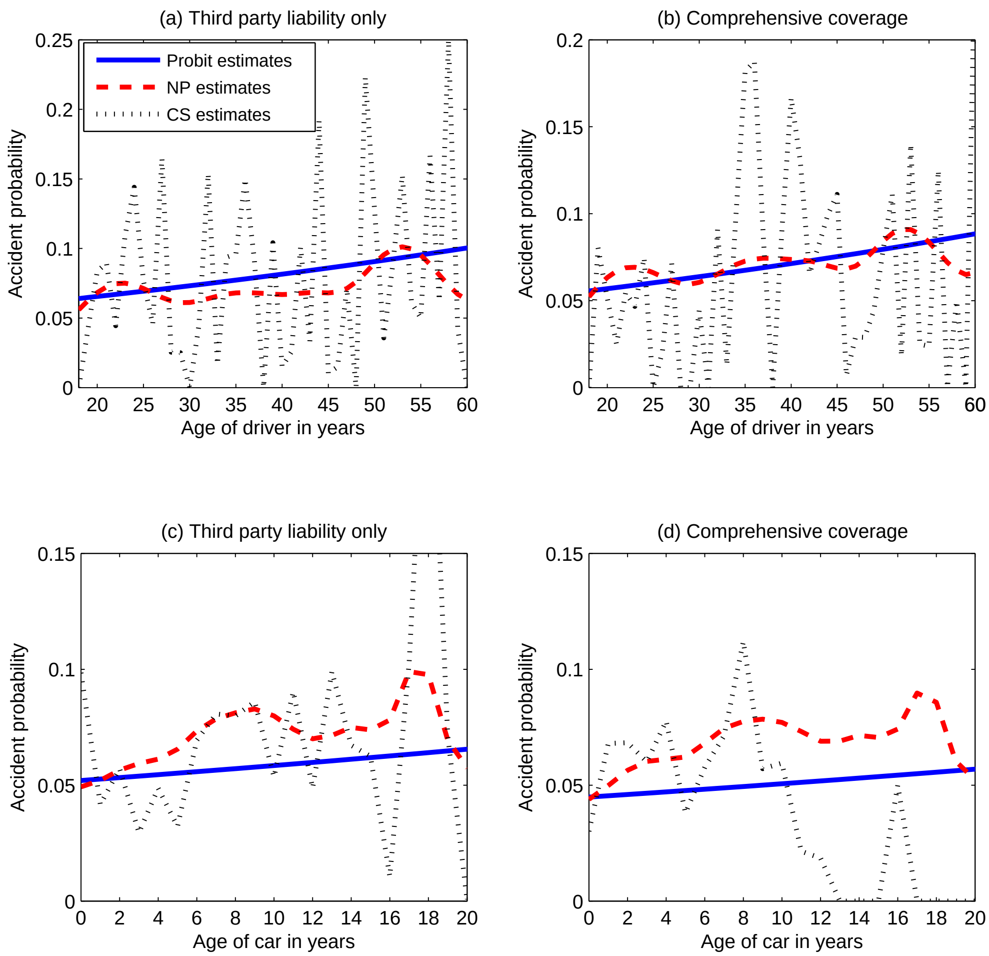

Figure 1: Predicted Accident Probabilities

Table 3: 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 the corresponding variable is treated as a continuous variable.

and the nonparametric method of Chiappori and Salanié (2000), respectively labeled as Probit estimates, NP estimates, and CS estimates in the figure. The third one is based on the empirical probabilities which are used for the nonparametric test by Chiappori and Salanié (2000). In order to calculate the accident probability for a certain age of driver, say, we restrict to all drivers with this age, classify them according to five binary risk variables and calculate for each cell (32 in total) the empirical accident probabilities. For each of these three methods the predicted probabilities are calculated as the averages of predictions over all observed values of the other variables in X_i. For example, when we calculate the predicted probabilities for the RC contract as a function of the age of driver in Figure 1, plot (a), the reported predicted probabilities are the averages of the estimates of

$$P\left(\text{At least one accident occurs} \mid \text{age of driver, other variables in } X_i,\ Y_i = 0\ \right)$$

as a function of the age of driver where the averages are taken over all observed values of the other variables in X_i.

We summarize some important findings from Figure 1. First, it shows that the probit model predicts a nearly linear monotonic relationship between the predicted (aggregate) accident probabilities and the age of driver (resp. age of car) under either contract, but our nonparametric estimates suggest that the relationship between the two may not be monotone in some region of the data. This indicates that the probit model might be inappropriate for the prediction of accident probabilities. Second, as expected, the predicted accident probabilities based on the CS method are not a smooth function of either continuous variable due to the transformation of all variables into binary variables and the fact that only information for a certain age of driver or car is used when we calculate the predicted accident probabilities for that age group. Third, when we focus on the estimated curves of predicted accident probabilities based on either the probit

Table 4: Bootstrap p values for our nonparametric test under various configurations

Configurations	I	II	III	IV	V	VI	VII	VIII	IX
Bootstrap p values	0.55	0.78	0.53	1.00	1.00	1.00	0.57	0.92	0.57

model or our nonparametric method, we find that the curves under the RC contract (plot (a) or (c) in the figure) share similar shapes as those under the TR contract (plot (b) or (d) in the figure), which suggests the lack of asymmetric information at the aggregate/average level.

Table 4 reports the bootstrap p-values for our nonparametric test under various configurations of the data set. We set the number of bootstrap replications B to 500. Table 4 suggests 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 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 Chiappori and Salanié's (2000) findings that there is no evidence of asymmetric information in the market for automobile insurance in France for young drivers. The results are very robust to different configurations of data.

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. Configurations VII-IX in Table 4 correspond to Configurations I-III with the only difference that the deductible now consists of three choices. Clearly, this modification 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 stake is not so high.

6.2 Long-Term Care Insurance

In this subsection we apply our nonparametric test to a long-term care insurance data set. The long-term care insurance (LTCI) covers an important risk which might lead to poverty among the elderly. Private information in the long-term care market in the US has been analyzed in the influential paper by Finkelstein and McGarry (2006). Below we first present a summary of Finkelstein and McGarry (2006), then describe their data set and finally present our empirical findings.

6.2.1 The Study of Finkelstein and McGarry (2006)

Finkelstein and McGarry (2006) find no positive correlation between individuals' insurance coverage and their risk experience even when controlling for the information which is given in the application form. Insurance coverage is a binary variable indicative of whether an individual had long-term care insurance in 1995. Risk is another binary variable indicative of whether an individual went into a nursing home in the five-year period between 1995 and 2000. Finkelstein and McGarry (2006) show that individuals have residual private information conditional on the risk assessment of the insurance company by using individuals' subjective assessments of the chance that they will enter a nursing home. They regress the variables insurance coverage and risk on the insurance companies' own assessment of the individuals' risk type (risk classification) in probit models and take up the subjective assessment as an additional exogenous variable. The subjective assessment is positively correlated with both risk and coverage when the risk classification of the insurance company is controlled for. They interpret this finding as a direct evidence of asymmetric information.

Finkelstein and McGarry (2006) argue that an explanation of the difference in the results is that individuals have private information about both their risk types and their preferences for insurance coverage. If individuals with private information that they have strong preferences for insurance are of lower risk, then private information about both risk types and preferences might operate in offsetting directions. This might finally result in a zero correlation between risk and coverage despite the presence of asymmetric information. Only through the use of an "unused observable" (information which is known to the insuree and the econometrician but not known or used by the insurance company for risk classification), namely, the subjective assessment mentioned above, could Finkelstein and McGarry (2006) find evidence of asymmetric information. But it is a very rare situation where such an "unused observable" is available. For this reason Gan, Huang, and Mayer (2011) propose a finite mixture model instead to account for differences in preferences for insurance and use it to analyze the data set of Finkelstein and McGarry (2006) without using any external information like the subjective assessment of individuals.

Table 5: Summary statistics for the long term care insurance data

Sample size	4,780
Number of insured persons	519 (10.8%)
Number of persons entering a nursing home	776 (16.2%)

Note: Percentages are in brackets.

6.2.2 Data Set and Risk Classification

Finkelstein and McGarry (2006) use individual-level survey data from the Asset and Health Dynamics (AHEAD) cohort of the Health and Retirement Study (HRS). Table 5 summarizes the key statistics of the data set. For more details about the data set we refer the readers directly to Finkelstein and McGarry (2006).

In order to conduct the positive correlation test, Finkelstein and McGarry (2006) use binary variables as proxy variables for risk and coverage as defined above. The data set contains similar detailed information which is used by insurance companies for risk classification: demographic information (age, gender, marital status, age of spouse), current health and medical history of the applicant. To replicate the information set of the insurer, Finkelstein and McGarry (2006) propose two approaches. First, they control for the insurance companies' actuarial prediction of the individual's risk type. This prediction is also used for pricing of the corresponding contract. In order to calculate this individual prediction of the probability that the insured will go into a nursing home over a five-year horizon, Finkelstein and McGarry (2006) apply the same actuarial model that is employed by insurance companies. The prediction depends nonparametrically on age, sex, and health state, measured by seven categories defined by the number of limitations to instrumental activities of daily living (IADLs), the number of limitations to activities of daily living (ADLs), and the absence of cognitive impairments. The model is described in Robinson (1996). In this case the exogenous variable X consists only of the insurance companies' individual risk prediction. In a second approach, Finkelstein and McGarry (2006) try to control for everything the insurance company observes about the individual. They include demographic information (age, marital status, age of spouse) and over 35 indicator variables for current health and health history. Moreover, they use income quartile, asset quartile and two-way and three-way interactions between certain selected variables.

We apply our nonparametric test to Finkelstein and McGarry's (2006) data set. As such a huge amount of indicator variables leads to a high number of cells and negligible number of observations per cell, we modify this setting and apply our test to two different configurations. We use the actuarial prediction for a claim, i.e., the use of long term care (LTC), as a continuous variable. Moreover, we use the number of ADLs, a proxy for current health and past health, sex and marriage status as discrete regressors. The proxy variable is a categorical variable that indicates the number of diseases/ limitations an individual suffers or suffered. Presumably these variables capture the decisive determinants on whether an elderly goes to a nursing home or not. We compile this data set as Configuration I and apply our nonparametric test. In Configuration

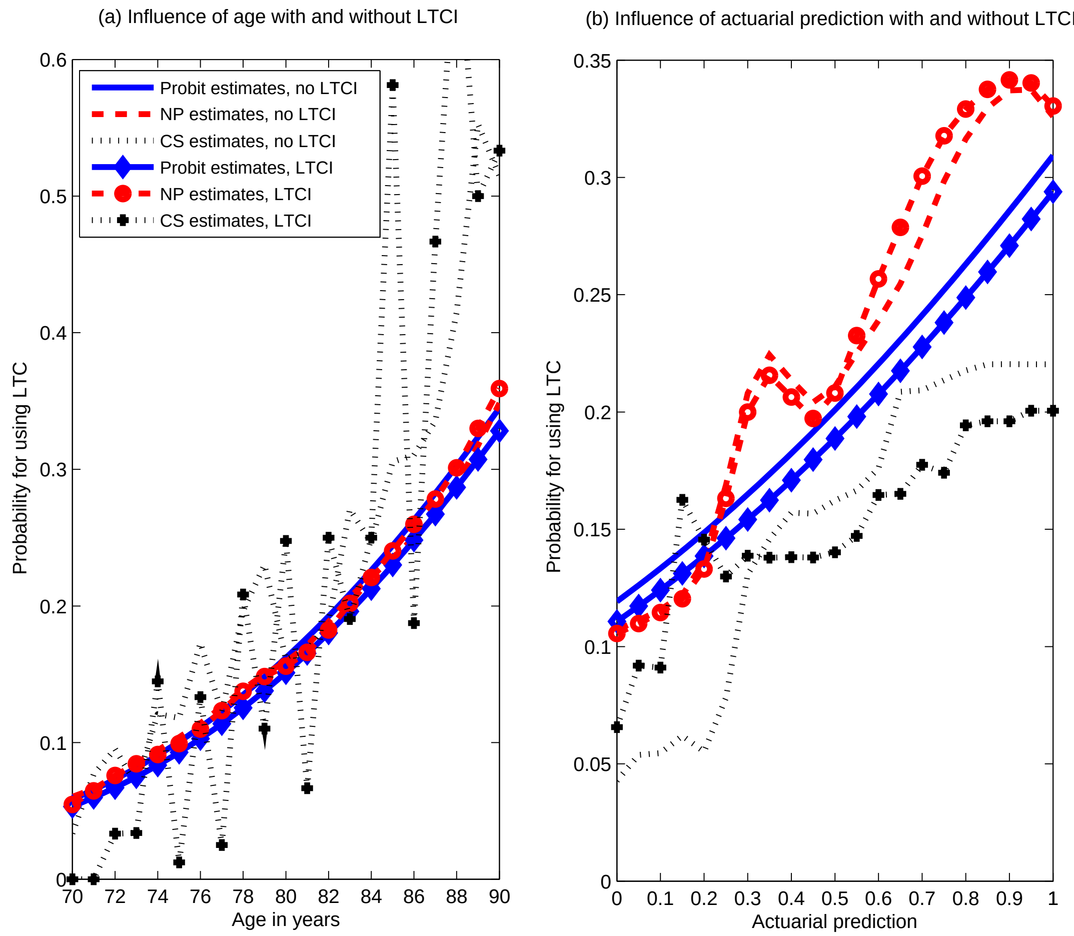

Figure 2: Predicted Probabilities of Entering a Nursing Home

II we additionally include the age of the insuree as a second continuous variable. The results are presented in the next section.

6.2.3 Empirical Results

In Figure 2 we plot the predicted probabilities of entering a nursing home over a five-year horizon as a function of the two continuous variables, namely, age and the actuarial prediction for the use of long term care, conditional on insurance coverage, i.e., with or without LTCI, by using all variables under Configuration II. As in the case of Figure 1, we calculate predicted probabilities based on a probit model, our nonparametric local constant estimate, and the nonparametric method of Chiappori and Salanié (2000), respectively labeled as Probit estimates, NP estimates, and CS estimates in Figure 2. In order to calculate the CS's predicted probabilities as a function of actuarial prediction that takes values between 0 and 1, we use a grid from 0 to 1 with step size 0.05 (with 21 points in total) to discretize the actuarial prediction. To appreciate the differences between the curves for individuals with and without LTCI, we follow the kind advice of a referee and lay the curves in the same plane.

We summarize some important findings from Figure 2. First, it shows that the probit model predicts a monotonic relationship between the predicted (aggregate) probabilities and age (resp. actuarial prediction) no matter whether an individual has LTCI or not, but our nonparametric estimates indicate the relationship between the predicted probabilities and age is monotonic but the predicted probabilities may depend on the actuarial prediction non-monotonically. Second, as expected, the predicted probabilities based on the CS's method are not a smooth function of age due to the transformation of all variables into binary variables and the fact that only information for a certain age of individuals is used when we calculate the predicted probabilities for that age group. But the predicted probabilities are a relatively smoother function of actuarial predication because we only use 21 points to discretize the latter. In any case, the CS's predictions are quite different from ours and those based on the probit model. Third, Figure 2(a) suggests that, both Probit and our NP methods yield similar estimates of the effects of age on the probability of entering a nursing home for the two groups of senior people with or without LTCI. Based on the Probit estimates, individuals with LTCI have lower probability of entering a nursing home than those without LTCI. But our NP estimates suggest that it is difficult to distinguish the estimated effects of age on the probability of entering a nursing home for individuals with or without LTCI. Fourth, Figure 2(b) suggests that our NP estimates are significantly different from the Probit estimates in terms of the effect of actuarial prediction on the probability of entering a nursing home. The Probit estimates of the effects of actuarial prediction on the probability of entering a nursing home are always higher for individuals without LTCI than for individuals with LTCI regardless the level of actuarial prediction. But this is not the case for our NP estimates. Fifth, our NP estimates reveal that the probability of entering a nursing home is lower for individuals with LTCI than those without LTCI when both have relatively low actuarial predictions $(0.2 - 0.5)$, and the probability of entering a nursing home is higher for those those with LTCI than those without LTCI when both have relatively high actuarial predictions $(0.5 - 1)$. Interestingly, for extremely low actuarial predictions $(0 - 0.2)$ the probabilities of entering a nursing home hardly differ for individuals with and without LTCI. In sum, we think our findings suggest some sort of asymmetric information that is related to risk preferences as opposed to risk types and thus lend support to Finkelstein and McGarry (2006).

For the two configurations described above, Table 6 reports the bootstrap p-values for our nonparametric test based on 500 bootstrap resamples. For either configuration, we can reject the null hypothesis at the 10% significance level. Therefore we conclude that there is some evidence of asymmetric information in the long-term care insurance market despite the fact that it is not overwhelmingly strong. The choice of contract contains information about the occurrence

Table 6: Bootstrap p values for our nonparametric test for the long term care insurance

Configurations	I	II
p value	0.060	0.082

of an accident. This result indicates that a functional and distributional specification is a very important issue in this field and that probit models might be inappropriate. In order to find some evidence for asymmetric information in the US long-term care insurance market, Finkelstein and McGarry (2006) have to resort to the so-called "unused observable". Our test works without the need to use such additional information which is only available in some exceptional cases.

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 consisting of novice drivers and to the US long term care insurance market. Our main conclusion for the car insurance data set is that we cannot detect asymmetric information in the data set despite different configurations of the control variables. 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 for young driver in the French car insurance. Our results are also very strong in contrast to Kim et al. (2009).

Our second application is the US long term care insurance market. While Finkelstein and McGarry (2006) find no positive correlation between risk and coverage – conditional on the variables used for risk classification – our test rejects the null hypothesis at the 10% significance level and therefore we detect the existence of asymmetric information in this market. Whereas Finkelstein and McGarry (2006) have to use additional information to establish their result, our test does not require additional information so that the test is widely applicable in most situations. Our analysis reveals also that a correct functional and distributional specification in this field is a very important issue and the use of probit models might be inappropriate. This application shows that our test may have power when other tests do not have.

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 or long term care 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.

Mathematical Appendix

A Proof of the Main Results

Let $\mathbf{B}_{ny}(\varsigma) \equiv E\{K_{h\lambda}(X_j, x)\mathbf{1}_j^y[F(z|X_j, Y_j) - F(z|x,y)]\}/f(x,y)$ and $\mathbf{V}_{ny}(\varsigma) \equiv n^{-1}\sum_{j=1}^n K_{h\lambda}$ $(X_j, x)\mathbf{1}_j^y[\mathbf{1}(Z_j \leq z) - F(z|X_j, Y_j)]/f(x,y)$. The following lemma establishes the uniform consistency of $\widehat{F}(z|x,y)$.

Lemma A.1 *Suppose Assumptions A1-A5 hold. Then uniformly in* $(x,y,z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ *we have:* $\widehat{F}(z|x,y) - F(z|x,y) = \mathbf{B}_{ny}(\varsigma) + \mathbf{V}_{ny}(\varsigma) + O_P((\nu_{1n} + \nu_{2n})^2) = O_P(\nu_{1n} + \nu_{2n})$, *where* $\nu_{1n} \equiv n^{-1/2}(h!)^{-1/2}\sqrt{\log n}$ *and* $\nu_{2n} \equiv \|h\|^2 + \|\lambda\|$.

Proof. Write $\widehat{F}(z|x,y) = \widehat{m}(z|x,y)/\widehat{f}(x,y)$ where $\widehat{m}(z|x,y) = \frac{1}{n}\sum_{j=1}^n K_{h\lambda}(X_j, x)\mathbf{1}_j^y\mathbf{1}(Z_j \leq z)$ and $\widehat{f}(x,y) = \frac{1}{n}\sum_{j=1}^n K_{h\lambda}(X_j, x)\mathbf{1}_j^y$. Then we have

$$\begin{aligned}\widehat{F}(z|x,y) - F(z|x,y) &= \frac{\widehat{m}(z|x,y) - \widehat{f}(x,y)F(z|x,y)}{f(x,y)}\left[1 + \frac{f(x,y) - \widehat{f}(x,y)}{\widehat{f}(x,y)}\right] \\ &= A_1(z|x,y) + A_2(z|x,y),\end{aligned}$$

where $A_1(z|x,y) = [\widehat{m}(z|x,y) - \widehat{f}(x,y)F(z|x,y)]/f(x,y)$ and $A_2(z|x,y) = [\widehat{m}(z|x,y) - \widehat{f}(x,y)$ $\times F(z|x,y)][f(x,y) - \widehat{f}(x,y)]/[\widehat{f}(x,y)f(x,y)]$. By Theorems 2, 7 and 8 in Hansen (2008) with little modification to account for discrete regressors, we can readily show that $A_2(z|x,y) = O_P((\nu_{1n} + \nu_{2n})^2)$ uniformly in $(x,y) \in \mathcal{X} \times \mathcal{Y}$. For $A_1(z|x,y)$, we have

$$\begin{aligned}A_1(z|x,y) &= \frac{1}{nf(x,y)}\sum_{j=1}^n K_{h\lambda}(X_j, x)\mathbf{1}_j^y[\mathbf{1}(Z_j \leq z) - F(z|x,y)] \\ &= \mathbf{V}_{ny}(\varsigma) + \mathbf{B}_{ny}(\varsigma) + \mathbf{R}_{ny}(\varsigma),\end{aligned}$$

where $\mathbf{R}_{ny}(\varsigma) = n^{-1}\sum_{j=1}^n\{K_{h\lambda}(X_j, x)\mathbf{1}_j^y[F(z|X_j, Y_j) - F(z|x,y)] - \mathbf{B}_{ny}(\varsigma)f(x,y)\}/f(x,y)$. By the same theorems, we can show that $\mathbf{V}_{ny}(\varsigma) = O_P(\nu_{1n})$, $\mathbf{B}_{ny}(\varsigma) = O(\nu_{2n})$, and $\mathbf{R}_{ny}(\varsigma) = O_P(\nu_{1n}\nu_{2n})$ uniformly in $(x,y) \in \mathcal{X} \times \mathcal{Y}$. It follows that $\widehat{F}(z|x,y) - F(z|x,y) = \mathbf{B}_{ny}(\varsigma) + \mathbf{V}_{ny}(\varsigma)$ $+ O_P((\nu_{1n} + \nu_{2n})^2) = O_P(\nu_{1n} + \nu_{2n})$ uniformly in $(x,y) \in \mathcal{X} \times \mathcal{Y}$. By the same argument as used in the proof of Theorem 4.1 of Boente and Fraiman (1991), we can show the above results also hold uniformly in $z \in \mathcal{Z}$ under Assumption A3. Then the conclusion follows. ■

Remark. Following Li and Racine (2008), we can write $\mathbf{B}_{ny}(\varsigma) = \sum_{s=1}^{p_c} h_s^2 B_{1s}(y|x,z) + \sum_{s=1}^{p_d}\lambda_s B_{2s}(y|x,z)$ +smaller order term, where $B_{1s}(z|x,y) \equiv \frac{1}{2}\int u^2 q(u)\,du[2F_s(z|x,y)f_s(x,y) + f(x,y)F_{ss}(z|x,y)]/f(x,y)$, $B_{2s}(z|x,y) \equiv \sum_{\widetilde{x}^d \in \mathcal{X}^d} I_s(\widetilde{x}^d, x^d)\left[F(z|x^c, \widetilde{x}^d, y)f(x^c, \widetilde{x}^d, y) - F(z|x,y)\right.$ $\left.f(x,y)\right]/f(x,y)$, $f_s(x,y) \equiv \partial f(x^c, x^d, y)/\partial x_s^c$, $F_s(z|x,y) \equiv \partial F(z|x^c, x^d, y)/\partial x_s^c$, $F_{ss}(z|x,y)$ $\equiv \partial^2 F(z|x^c, x^d, y)/\partial(x_s^c)^2$, and $I_s(\widetilde{x}^d, x^d) = \mathbf{1}(x_s^d \neq \widetilde{x}_s^d)\Pi_{r=1,r\neq s}^{p_d}\mathbf{1}(x_r^d = \widetilde{x}_r^d)$.

Proof of Theorems 4.1 and 4.2

We only prove Theorem 4.2, as the proof of Theorem 4.1 is a special case.

First, we observe that $(h!)^{1/2} D_n = (h!)^{1/2} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \sum_{i=1}^{n} [\widehat{F}(Z_i|X_i, r) - \widehat{F}(Z_i|X_i, s)]^2 a_i$ $= D_{n1} + D_{n2} + 2D_{n3}$ where $a_i \equiv a(X_i^c)$,

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

Under $\mathbb{H}_1(n^{-1/2}(h!)^{-1/4})$, we prove the theorem by showing that (i) $D_{n1} \xrightarrow{P} \mu_0$, (ii) $D_{n2} - B_0 \xrightarrow{d} N(0, \sigma_0^2)$, (iii) $D_{n3} = o_P(1)$, (iv) $\widehat{B}_n = B_0 + o_P(1)$, 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 $\mathbb{H}_1(n^{-1/2}(h!)^{-1/4})$. It remains to show (ii)-(iv).

To show (ii), we first apply Lemma A.1 to obtain

$$\begin{aligned} D_{n2} &= (h!)^{1/2} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \sum_{i=1}^{n} \left\{[\mathbf{V}_{nr}(\varsigma_i) - \mathbf{V}_{ns}(\varsigma_i)] + [\mathbf{B}_{nr}(\varsigma_i) - \mathbf{B}_{ns}(\varsigma_i)] + O_P(\nu_{1n}^2 + \nu_{2n}^2)\right\}^2 a_i \\ &= (h!)^{1/2} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \sum_{i=1}^{n} [\mathbf{V}_{nr}(\varsigma_i) - \mathbf{V}_{ns}(\varsigma_i)]^2 a_i \\ &\quad + 2(h!)^{1/2} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \sum_{i=1}^{n} [\mathbf{V}_{nr}(\varsigma_i) - \mathbf{V}_{ns}(\varsigma_i)] [\mathbf{B}_{nr}(\varsigma_i) - \mathbf{B}_{ns}(\varsigma_i)] a_i \\ &\quad + (h!)^{1/2} \sum_{i=1}^{n} [\mathbf{B}_{nr}(\varsigma_i) - \mathbf{B}_{ns}(\varsigma_i)]^2 a_i + n(h!)^{1/2} O_P(\nu_{1n}^3 + \nu_{2n}^3) \\ &\equiv D_{n21} + 2D_{n22} + D_{n23} + o_P(1) \end{aligned} \tag{A.1}$$

where the definitions of D_{n21}, D_{n22}, and D_{n23} are self-evident. Let $v_{ny}(\xi_j, \varsigma) \equiv K_{h\lambda}(X_j, x) \mathbf{1}_j^y$ $\times [\mathbf{1}(Z_j \leq z) - F(z|X_j, Y_j)] / f(x, y)$ and $\varphi_{rs}(\xi_i, \xi_j) \equiv v_{nr}(\xi_j, \varsigma_i) - v_{ns}(\xi_j, \varsigma_i)$. Noting that $\mathbf{V}_{ny}(\varsigma) = n^{-1} \sum_{j=1}^{n} v_{ny}(\xi_j, \varsigma)$ and $\mathbf{V}_{nr}(\varsigma_i) - \mathbf{V}_{ns}(\varsigma_i) = n^{-1} \sum_{j=1}^{n} \varphi_{rs}(\xi_i, \xi_j)$, we have $D_{n21} = \frac{(h!)^{1/2}}{n^2} \sum_{i=1}^{n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} [\sum_{j=1}^{n} \varphi_{rs}(\xi_i, \xi_j)]^2 a_i = V_n + B_n + R_n$, where $V_n \equiv \frac{(h!)^{1/2}}{n^2} \sum_{i=1}^{n} \sum_{j \neq i}^{n}$ $\sum_{k \neq i,j}^{n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \varphi_{rs}(\xi_i, \xi_j) \varphi_{rs}(\xi_i, \xi_k) a_i$, $B_n \equiv \frac{(h!)^{1/2}}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \varphi_{rs}(\xi_i, \xi_j)^2$ $\times a_i$, and $R_n \equiv \frac{2(h!)^{1/2}}{n^2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \varphi_{rs}(\xi_i, \xi_j) \varphi_{rs}(\xi_i, \xi_i) a_i$. Let $\overline{\varphi}_{rs}(\xi_i, \xi_j, \xi_k) \equiv$ $[\varphi_{rs}(\xi_i, \xi_j) \varphi_{rs}(\xi_i, \xi_k) a_i + \varphi_{rs}(\xi_j, \xi_i) \varphi_{rs}(\xi_j, \xi_k) a_j + \varphi_{rs}(\xi_k, \xi_i) \varphi_{rs}(\xi_k, \xi_j) a_k] / 3$. 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 $\overline{V}_n \equiv \frac{6(h!)^{1/2}}{n(n-1)(n-2)} \sum_{1\le i<j<k\le n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \overline{\varphi}_{rs}\left(\xi_i, \xi_j, \xi_k\right)$. Note that for all $i \neq j \neq k$, $\theta \equiv E\left[\overline{\varphi}_{rs}\left(\xi_i, \xi_j, \xi_k\right)\right] = 0$, $\overline{\varphi}_{rs,1}(b) \equiv E\left[\overline{\varphi}_{rs}\left(b, \xi_j, \xi_k\right)\right] = 0$, and $\overline{\varphi}_{rs,2}(b, \widetilde{b}) \equiv E[\overline{\varphi}_{rs}(b, \widetilde{b}, \xi_k)] = \frac{1}{3}E[\varphi_{rs}(\xi_k, b)\ \varphi_{rs}(\xi_k, \widetilde{b})]$. Let $\overline{\varphi}_{rs,3}(b, \widetilde{b}, \overline{b}) \equiv \overline{\varphi}_{rs}(b, \widetilde{b}, b) - \overline{\varphi}_{rs,2}(b, \widetilde{b}) - \overline{\varphi}_{rs,2}(b, \overline{b}) - \overline{\varphi}_{rs,2}(\widetilde{b}, b)$. By the Hoeffding decomposition,

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

where $H_n^{(2)} \equiv \frac{2(h!)^{1/2}}{n(n-1)} \sum_{1\le i<j\le n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \overline{\varphi}_{rs,2}\left(\xi_i, \xi_j\right)$ and $H_n^{(3)} \equiv \frac{6(h!)^{1/2}}{n(n-1)(n-2)} \sum_{1\le i<j<k\le n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \overline{\varphi}_{rs,3}(\xi_i, \xi_j, \xi_k)$. Noting that $E\left[\overline{\varphi}_{rs,3}\left(a, \widetilde{a}, \xi_i\right)\right] = 0$ 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, $H_n^{(3)} = O_P(n^{-3/2}(h!)^{-1/2}) = o_P\left(n^{-1}\right)$ 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

$$\begin{aligned}
\mathcal{H}_n &\equiv \frac{2(h!)^{1/2}}{n} \sum_{1\le i<j\le n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} 3\overline{\varphi}_{rs,2}\left(\xi_i, \xi_j\right) \\
&= \frac{2(h!)^{1/2}}{n} \sum_{1\le i<j\le n} \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} \int \varphi_{rs}\left(\xi, \xi_i\right) \varphi_{rs}\left(\xi, \xi_j\right) a\left(x^c\right) F_\xi\left(d\xi\right).
\end{aligned}$$

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 $\mathcal{H}_n : \mathcal{H}_n \xrightarrow{d} N\left(0, \lim_{n\to\infty}\sigma_n^2\right)$, where $\sigma_n^2 \equiv 2h!E_iE_j[\sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1}\int \varphi_{rs}\left(\xi, \xi_i\right)\varphi_{rs}\left(\xi, \xi_j\right) \times a\left(x^c\right) F_\xi\left(d\xi\right)]^2 = 2h!E_iE_j\{\sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1}\int\left[v_{nr}\left(\xi_i, \varsigma\right) - v_{ns}\left(\xi_i, \varsigma\right)\right]\left[v_{nr}\left(\xi_j, \varsigma\right) - v_{ns}\left(\xi_j, \varsigma\right)\right] a\left(x^c\right) F_\xi\left(d\xi\right)\}^2$, where E_i denotes expectation with respect to ξ_i. We now show that $\lim_{n\to\infty}\sigma_n^2 = \sigma_0^2$ which is defined in (4.9). In view of the fact that $v_{nr}\left(\xi_i, \varsigma\right) v_{ns}\left(\xi_i, \varsigma\right) = 0$ for any $r \neq s$, we have

$$\begin{aligned}
\sigma_n^2 &= 2h!E_iE_j\left\{\sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1}\int[v_{nr}\left(\xi_i, \varsigma\right) v_{nr}\left(\xi_j, \varsigma\right) + v_{ns}\left(\xi_i, \varsigma\right) v_{ns}\left(\xi_j, \varsigma\right)]a\left(x^c\right) F_\xi\left(d\xi\right)\right\}^2 \\
&= 2h!E_iE_j\left\{\sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1}\overline{v}_{nr}(\xi_i, \xi_j)\right\}^2 + 2h!E_iE_j\left\{\sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1}\overline{v}_{ns}(\xi_i, \xi_j)\right\}^2 \\
&\quad +4h!E_iE_j\left[\sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1}\overline{v}_{nr}(\xi_i, \xi_j)\sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1}\overline{v}_{ns}(\xi_i, \xi_j)\right] \\
&\equiv \sigma_{1n}^2 + \sigma_{2n}^2 + \sigma_{3n}^2,\ \text{say},
\end{aligned}$$

where $\overline{v}_{nr}(\xi_i, \xi_j) \equiv \int v_{nr}\left(\xi_i, \varsigma\right) v_{nr}\left(\xi_j, \varsigma\right) a\left(x^c\right) F_\xi\left(d\xi\right)$. By straightforward but tedious calculations, one can verify that $\sigma_{1n}^2 = 2C_2\sum_{r=0}^{c_y-2}\left(c_y - 1 - r\right)^2\alpha_r + o(1)$, $\sigma_{2n}^2 = 2C_2\sum_{r=1}^{c_y-1} r^2\alpha_r + o(1)$, and $\sigma_{3n}^2 = 4C_2\sum_{r=1}^{c_y-2}\left(c_y - 1 - r\right) r\alpha_r + o(1)$, where $\alpha_r = \sum_{x^d\in\mathcal{X}^d}\int\int\int f\left(x, r\right)^{-2} V\left(z, \overline{z}; x, r\right)^2 f\left(x\right) dF\left(z|x\right) dF\left(\overline{z}|x\right) a\left(x^c\right)^2 dx^c$. It follows that $\sigma_n^2 = 2C_2\left(c_y - 1\right)^2\sum_{r=0}^{c_y-2}\alpha_r + o(1) = \sigma_0^2 + o(1)$. Consequently $V_n \xrightarrow{d} N\left(0, \sigma_0^2\right)$. By moment calculations, the Chebyshev inequality, and Assumptions

A1 and A4-A5, we can show that

$$\begin{aligned} B_n &= \frac{(h!)^{1/2}}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1}\left\{\left[v_{nr}\left(\xi_j,\varsigma_i\right)\right]^2+\left[v_{ns}\left(\xi_j,\varsigma\right)\right]^2\right\}a\left(X_i^c\right) \\ &= \frac{(h!)^{1/2}\left(c_y-1\right)}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{r=0}^{c_y-1}\left[v_{nr}\left(\xi_j,\varsigma_i\right)\right]^2 a\left(X_i^c\right)=B_0+o_P\left(1\right). \end{aligned}$$

For R_n, it is easy to verify that $E\left(R_n\right)=0$ and $E\left(R_n^2\right)=O\left((nh!)^{-1}\right)=o\left(1\right)$. So $R_n=o_P\left(1\right)$ by the Chebyshev inequality. Consequently, we have

$$D_{n21}-B_n \xrightarrow{d} N\left(0,\sigma_0^2\right). \tag{A.2}$$

Let $b_{rs}\left(\varsigma_i\right)\equiv \mathbf{B}_{nr}\left(\varsigma_i\right)-\mathbf{B}_{ns}\left(\varsigma_i\right)$. Then $D_{n22}=\sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1}(D_{n22,rs1}-D_{n22,rs2})$, where $D_{n22,rs1}\equiv(h!)^{1/2}\sum_{i=1}^{n}\mathbf{V}_{nr}\left(\varsigma_i\right)b_{rs}\left(\varsigma_i\right)$ and $D_{n22,rs2}\equiv(h!)^{1/2}\sum_{i=1}^{n}\mathbf{V}_{ns}\left(\varsigma_i\right)b_{rs}\left(\varsigma_i\right)$. Write

$$\begin{aligned} D_{n22,rs1} &= n^{-1}\left(h!\right)^{1/2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} f\left(X_i,r\right)^{-1}K_{h\lambda}\left(X_j,X_i\right)\mathbf{1}_j^r\left[\mathbf{1}\left(Z_j\leq Z_i\right)-F\left(Z_i|X_j,Y_j\right)\right]b_{rs}\left(\varsigma_i\right)a_i \\ &\quad +n^{-1}\left(h!\right)^{1/2}\sum_{i=1}^{n} f\left(X_i,r\right)^{-1}K_{h\lambda}\left(X_i,X_i\right)\mathbf{1}_i^r\left[\mathbf{1}\left(Z_i\leq Z_i\right)-F\left(Z_i|X_i,Y_i\right)\right]b_{rs}\left(\varsigma_i\right)a_i \\ &\equiv D_{n22,rs1a}+D_{n22,rs1b},\ \text{say.} \end{aligned}$$

Observing that $b_{rs}\left(\varsigma_i\right)=O(v_{2n})$, it is straightforward to show that $D_{n22,rs1b}=O_P((h!)^{-1/2}\,v_{2n})=o_P\left(1\right)$ under Assumption A5. Noting that $E\left(D_{n22,rs1a}\right)=0$ and $E\left(D_{n22,rs1a}^2\right)=O(nh!v_{2n}^2)=o\left(1\right)$, we have $D_{n22,rs1a}=o_P\left(1\right)$ by the Chebyshev inequality. Hence $D_{n22,rs1}=o_P\left(1\right)$. By the same token, $D_{n22,rs2}=o_P\left(1\right)$. It follows that

$$D_{n22}=o_P\left(1\right). \tag{A.3}$$

By Lemma A.1 and Assumption A5, we have $D_{n23}=n\left(h!\right)^{1/2}O_P(v_{2n}^2)=o_P\left(1\right)$. This, in conjunction with (A.1), (A.2) and (A.3), implies that $D_{n2}-B_n\xrightarrow{d}N\left(0,\sigma_0^2\right)$.

Next, we show (iii). By Lemma A.1, under $\mathbb{H}_1\left(n^{-1/2}(h!)^{-1/4}\right)$ we have $D_{n3}=\sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1}$ $(D_{n31,rs}+D_{n32,rs})+n^{1/2}\left(h!\right)^{1/4}O_P\left(\nu_{1n}^2+\nu_{2n}^2\right)=\sum_{r=0}^{c_y-2}\sum_{s=r+1}^{c_y-1}\left(D_{n31,rs}+D_{n32,rs}\right)+o_P\left(1\right)$, where $D_{n31,rs}\equiv n^{-1/2}\left(h!\right)^{1/4}\sum_{i=1}^{n}[\mathbf{V}_{nr}\left(\varsigma_i\right)-\mathbf{V}_{ns}\left(\varsigma_i\right)]\delta_{n,rs}(\varsigma_i)a_i$ and $D_{n32,rs}\equiv n^{-1/2}\left(h!\right)^{1/4}\sum_{i=1}^{n}$ $[\mathbf{B}_{nr}\left(\varsigma_i\right)-\mathbf{B}_{ns}\left(\varsigma_i\right)]\delta_{n,rs}(\varsigma_i)a_i$. As in the analysis of $D_{n22,rs}$, we can readily show that $D_{n31,rs}=O_P(h!^{1/2}+n^{-1/2}\left(h!\right)^{-3/4})=o_P\left(1\right)$ by Assumption A5. Noting that $\sup_{\varsigma}\left|\mathbf{B}_{nr}\left(\varsigma\right)\right|=O(v_{2n})$, we have $D_{n32,rs}\leq n^{1/2}\left(h!\right)^{1/4}O\left(v_{2n}\right)n^{-1}\sum_{i=1}^{n}\left|\delta_{n,rs}(\varsigma_i)\right|=O_P(n^{1/2}\left\|h\right\|^2\left(h!\right)^{1/4})=o_P\left(1\right)$. Consequently, $D_{n3}=o_P\left(1\right)$.

We now show (iv). Let $d_1 \equiv C_1 (h!)^{-1/2} (c_y - 1)$. Then we have

$$\begin{aligned}
\widehat{B}_n - B_0 &= \frac{d_1}{n} \sum_{i=1}^{n} \sum_{r=0}^{c_y-1} \left[\widehat{f}(X_i, r)^{-1} \widehat{\sigma}^2 (Z_i|X_i, r) - f(X_i, r)^{-1} \sigma^2 (Z_i|X_i, r) \right] a(X_i^c) \\
&\quad + \frac{d_1}{n} \sum_{i=1}^{n} \sum_{r=0}^{c_y-1} \left\{ f(X_i, r)^{-1} \sigma^2 (Z_i|X_i, r) a(X_i^c) - E[f(X_i, r)^{-1} \sigma^2 (Z_i|X_i, r) a(X_i^c)] \right\} \\
&\equiv B_{1n} + B_{2n}, \text{ say.}
\end{aligned}$$

By the Chebyshev inequality, it is straightforward to show that $B_{2n} = O_P((nh!)^{-1/2}) = o_P(1)$. Use the fact that $\widehat{f}(x, r) - f(x, r) = O_P(v_{1n} + v_{2n})$ and $\widehat{\sigma}^2(z|x, r) - \sigma^2(z|x, r) = O_P(v_{1n} + v_{2n})$ uniformly in $(x, r, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, we can readily show that $B_{1n} = (h!)^{-1/2} O_P(v_{1n} + v_{2n}) = o_P(1)$ under Assumption A5. Consequently, $\widehat{B}_n - B_0 = o_P(1)$.

For (v), letting $d_2 \equiv 2C_2 (c_y - 1)^2$ we have

$$\begin{aligned}
&\widehat{\sigma}_n^2 - \sigma_0^2 \\
&= \frac{d_2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{r=0}^{c_y-1} \left[\widehat{\beta}_r(\varsigma_i, \varsigma_j) - \beta_r(\varsigma_i, \varsigma_j) \right] K_{h\lambda, ji} a(X_i^c) a(X_j^c) \\
&\quad + \frac{d_2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{r=0}^{c_y-1} \left\{ \beta_r(\varsigma_i, \varsigma_j) K_{h\lambda, ji} a(X_i^c) a(X_j^c) - E\left[\beta_r(\varsigma_i, \varsigma_j) K_{h\lambda, ji} a(X_i^c) a(X_j^c) \right] \right\} \\
&\quad + \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \left\{ d_2 \sum_{r=0}^{c_y-1} E\left[\beta_r(\varsigma_i, \varsigma_j) K_{h\lambda, ji} a(X_i^c) a(X_j^c) \right] - \sigma_0^2 \right\} \\
&\equiv C_{1n} + C_{2n} + C_{3n}, \text{ say,}
\end{aligned}$$

where $\beta_r(\varsigma_i, \varsigma_j) = f(X_i, r)^{-1} f(X_j, r)^{-1} V(Z_i, Z_j; X_i, r) V(Z_i, Z_j; X_j, r)$ and $\widehat{\beta}_r(\varsigma_i, \varsigma_j) = \widehat{f}(X_i, r)^{-1}$ $\widehat{f}(X_j, r)^{-1} \widehat{V}(Z_i, Z_j; X_i, r) \widehat{V}(Z_i, Z_j; X_j, r)$. Using the uniform consistency of $\widehat{f}(x, r)$ and $\widehat{F}(y|x, r)$, we can readily show that $C_{1n} = O_P(v_{1n} + v_{2n}) = o_P(1)$. By the law of large numbers for second order U-statistic, $C_{2n} = o_P(1)$. By moment calculations, $C_{3n} = O(v_{2n}) = o(1)$. Thus we have $\widehat{\sigma}_n^2 = \sigma_0^2 + o_p(1)$. ■

Proof of Theorems 4.3

Using the notation defined in the proof of Theorem 4.2, we have $n^{-1} D_n = n^{-1} (h!)^{-1/2} (D_{n1} + D_{n2} + 2D_{n3})$. Under $\mathbb{H}_1$, it is easy to show that $n^{-1} (h!)^{-1/2} D_{n1} = \sum_{r=0}^{c_y-2} \sum_{s=r+1}^{c_y-1} E[F(Z_i|X_i, r) - F(Z_i|X_i, s)]^2 + o_P(1)$, $n^{-1} (h!)^{-1/2} D_{n2} = O_P(\nu_{1n}^2 + \nu_{2n}^2) = o_P(1)$, and $n^{-1} (h!)^{-1/2} D_{n3} = O_P(\nu_{1n} + \nu_{2n}) = o_P(1)$. 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. ■

ACKNOWLEDGMENTS

The authors sincerely thank Keisuke Hirano, the associate editor, and three anonymous referees for their many insightful comments and suggestions that lead to a substantial improvement of the presentation. They are very grateful to Bernard Salanié for his most valuable comments and for providing us with the first data set used in the paper. They gratefully acknowledge a referee for pointing to them that Dionne, Gourieroux, and Vanasse (2001, 2006) first formulate the test of asymmetric information as a test of conditional independence as outlined in Section 3.1. Spindler gratefully acknowledges financial support from the Deutsche Forschungsgemeinschaft through GRK 801.

References

Aitchison, J., and 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., and 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., and Heckman, J. J. (1999), "Testing for Moral Hazard Using Dynamic Data," *Manuscript*, Dept. of Economics, University of Chicago.

Chiappori, P.-A., and Salanié, B. (1997), "Empirical Contract Theory: the Case of Insurance Data," *European Economic Review* 41, 943-950.

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

Chiappori, P.-A., Jullien, B., Salanié, B., and 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," *Review of Economics and Statistics* 87, 197-207.

Cohen, A., and Siegelman, P. (2010), "Testing for Adverse Selection in Insurance Markets," *Journal of Risk and Insurance* 77, 39-84.

Crocker, K.-J., and Snow, A. (2000), "The Theory of Risk Classification," in *Handbook of Insurance*, eds., G. Dionne, pp. 245-276. Kluwer Academic Publishers.

De Jong, P. and Heller, G. Z. (2008), *Generalized Linear Models for Insurance Data,* Oxford: Oxford University Press.

Delgado, M. A., and González-Manteiga, W. (2001), "Significance Testing in Nonparametric Regression Based on the Bootstrap," *Annals of Statistics* 29, 1469-1507.

Dionne, G., Doherty, N., and Fombaron, N. (2000), "Adverse Selection in Insurance Markets," in *Handbook of Insurance*, eds. G. Dionne, pp. 185-244. Kluwer Academic Publishers.

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

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

Dionne, G., Michaud, P.-C., and Dahchour, M. (2012), "Separating Moral Hazard from Adverse Selection in Automobile Insurance: Longitudinal Evidence from France," *Journal of the European Economic Association* forthcoming.

Finkelstein, A., and McGarry, K. (2006), "Multiple Dimensions of Private Information: Evidence from the Long-term Care Insurance Market," *American Economic Review* 96, 938-958.

Finkelstein, A., and Poterba, J. (2006), "Testing for Adverse Selection with 'Unused Observables'," *NBER Working Paper* 12112.

Gan, L., Huang, F., and Mayer, A. (2011), "A Simple Test of Private Information in the Insurance Markets with Heterogeneous Insurance Demand," *NBER Working Paper* 16738.

Gourieroux, C., Monfort, A., Renault, E., and 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.

Horowitz, J. L., and Spokoiny, V. G. (2001), "An Adaptive, Rate-optimal Test of a Parametric Mean-regression Model against a Nonparametric Alternative," *Econometrica,* 69, 599-631.

Huang, M., and White, H. (2009), "A flexible Test for Conditional Independence," *Working Paper*, Dept. of Economics, UCSD.

Huang, T.-M. (2010), "Testing Conditional Independence Using Maximal Nonlinear Conditional Independence," *Annals of Statistics,* 38, 2047-2091.

Jullien, B., Salanié, B., and Salanié, F. (2007), "Screening Risk Averse Agents under Moral Hazard," *Economic Theory* 30, 151-169.

Kaas, R., Goovaerts, M., Dhaene, J., and Denuit, M. (2008), *Modern Actuarial Risk Theory: Using R*, Springer, 2nd edition.

Kim, H., Kim, D., Im, S., and Hardin, J. W. (2009), "Evidence of Asymmetric Information in the Automobile Insurance Market: Dichotomous versus Multinomial Measurement of Insurance Coverage," *Journal of Risk and Insurance* 76, 343-366.

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

Li, Q., and 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.

Mikosch, T. (2006), *Non-Life Insurance Mathematics: An Introduction with Stochastic Processes.* Springer.

Paparoditis, E., and 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., and Snow, A. (1994), "Evidence on Adverse Selection: Equilibrium Signaling and Cross-subsidization in the Insurance Market," *Journal of Political Economy* 102, 236-257.

Racine, J., and 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, J. (1996), "A Long-term Care Status Transition Model," presented at Bowles Symposium on the Old-Age Crisis: Actuarial Opportunities, Georgia State University, September 26-27, 1996.

Rothschild, M., and Stiglitz, J. E. (1976) "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *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., and Ullah, A. (2009), "Testing Conditional Uncorrelatedness," *Journal of Business and Economic Statistics,* 27, 18-29.

Su, L., and White, H. (2007), "A Consistent Characteristic-function-based Test for Conditional Independence," *Journal of Econometrics* 141, 807-834.

Su, L., and White, H. (2008), "A Nonparametric Hellinger Metric Test for Conditional Independence," *Econometric Theory* 24, 829-864.

Winter, R. (2000), "Optimal Insurance under Moral Hazard," in *Handbook of Insurance*, eds. G. Dionne, pp.155-184. Kluwer Academic Publishers.