North American Actuarial Journal

ISSN: 1092-0277 (Print) 2325-0453 (Online) Journal homepage: http://www.tandfonline.com/loi/uaaj20

Managed Care and Health Care Utilization: Specification of Bivariate Models Using Copulas

Peng Shi & Wei Zhang

To cite this article: Peng Shi & Wei Zhang (2013) Managed Care and Health Care Utilization: Specification of Bivariate Models Using Copulas, North American Actuarial Journal, 17:4, 306-324, DOI: 10.1080/10920277.2013.849192

To link to this article: https://doi.org/10.1080/10920277.2013.849192

Published online: 26 Nov 2013.

North American Actuarial Journal, 17(4), 306–324, 2013

ISSN: 1092-0277 print / 2325-0453 online
DOI: 10.1080/10920277.2013.849192

Routledge
Taylor & Francis Group

Managed Care and Health Care Utilization: Specification of Bivariate Models Using Copulas

Peng Shi[1] and Wei Zhang[2]
[1]*Department of Actuarial Science, Risk Management, and Insurance, Wisconsin School of Business, University of Wisconsin–Madison, Madison, Wisconsin*
[2]*Department of Economics, Northern Illinois University, DeKalb, Illinois*

This article studies the effect of managed care on health care utilization compared to traditional fee-for-service plans in private health insurance market. To construct our hypothesis, we build a game-theoretic model to study health care utilization under a two-sided moral hazard: of patients and providers. In econometric modeling, we employ a copula regression to jointly examine individuals' health plan choice and their utilization of medical care services, because of the endogeneity of insurance choice. The dependence parameter in the copula reflects the relation between the two outcomes, based on which the average treatment effects are further derived. We apply the methodology to a survey data set of the U.S. population and consider three types of curative care and three types of preventive care for the measurement of medical care utilization. We find that managed care is in general associated with higher care utilization. Evidence is also found on the underlying incentives of both patients and medical providers.

1. INTRODUCTION

Health care expenditures account for a substantive proportion of gross domestic product (GDP) in all OECD countries and the percentage has trended upward over the past decade. In the United States expenditures for medical care services, as a share of GDP, have increased from 14% in 2000 to 18% in 2009, according to the *2012 Statistical Abstract of the United States* (Executive Office of the President 2009). High health care costs to some extent reflect inefficiency in the current health services delivery system. According to *The Economic Case for Health Care Reform*, a special report by the Council of Economic Advisers (CEA), one important cause of such inefficiency is the information asymmetry among the three key players in the market; patients, medical care providers, and insurance companies. To align the incentives of market participants, different types of health plans have been designed and implemented to deliver medical care services. Most existing plans are either managed care or traditional fee-for-service plans.

In this study we compare the utilization of medical care services under different types of health plans. In particular, we are interested in whether the widely used managed care increases or decreases utilization when compared to traditional fee-for-service plans. What complicates the investigation are the offsetting effects of the incentives of patients and care providers on care utilization and the interactive relation between health care utilization and choice of health plan. The two factors are separate yet related. The incentives from both patients and physicians affect the ultimate care utilization and the selection of health insurance, though in a different fashion.

Health care utilization has been an important topic in the health economics literature (see, for example, Jiménez-Martín et al. 2002; d'Uva 2006). Even in the context of current debates on health care reform, the initiation of an expanded national program of care motivates us to better understand how health care costs are influenced by various factors, among which the design of the health plan is of particular importance from the perspective of improving the efficiency of the health care delivery system. Historically, managed care plans seek to deliver cost-effective care, and the effects on health care utilization has been investigated extensively; see Miller and Luft (1994) and Glied (2000) for a review. However, much of the research has neglected the endogeneity of insurance selection; that is, the utilization of health care and the choice of health plan are simultaneously determined by individuals. In

Financial support from the Actuarial Foundation and the Society of Actuaries is gratefully acknowledged. We thank the Editor, Mary Hardy, the Associate Editor, and two anonymous reviewers for helpful comments and suggestions.
Address correspondence to Peng Shi, Department of Actuarial Science, Risk Management, and Insurance, Wisconsin School of Business, University of Wisconsin–Madison, 5281 Grainger Hall, 975 University Avenue, Madison, WI 53706. E-mail: pshi@bus.wisc.edu

recent work, Deb et al. (2006) and Deb and Trivedi (2006) used latent factors to account for selection on unobservables, where a computationally expensive simulation-based likelihood method is required for estimation.

To address the simultaneity issue, we take a full information approach, by using copulas to jointly model health plan choice and medical care usage. Another appealing aspect of the copula model is that it allows us to identify the incentives from both the patient and the physician side. Our analysis is built on the hypothesis that incentives from patients and physicians will lead to an opposite relation between care utilization and plan types, and the resulting relation could go either way. This seemingly intuitive observation is not trivial, because the ultimate care utilization is determined by the interactive decision process of patients and physicians through the incentive provided by health plan. To provide a rigorous foundation for the empirical analysis, we develop a game-theoretical model to study patient and provider behaviors in an integrated manner, and we show that the equilibrium may display either positive or negative relation between care utilization and plan types.

In the empirical analysis, we consider a sample of the U.S. civilian noninstitutionalized population from the Medical Expenditure Panel Survey (MEPS). The MEPS data have been used in the literature of health economics (see, for example, Deb et al. 2006; Zimmer and Trivedi 2006) and actuarial studies (see, for example, Frees 2010; Frees et al. 2011). More detailed information on MEPS is discussed in Section 4. Two types of care are considered: preventive care and curative care. Using the definition in MEPS data, the usage of preventive care is measured by whether a certain type of preventive care is used, and the usage of curative care is measured by the number of visits to medical providers to seek such care. We find that managed care in general is associated with higher utilization of care after purging off individual heterogeneity. To further quantify the difference of care utilization under different types of plans, we demonstrate the calculation of the treatment effect from the copula model for an individual with average characteristics.

The dependence parameter in the copula captures a residual relation between care utilization and plan type, and it represents the offsetting effects of incentives from patients and physicians. The significant positive association supports only the incentives from the patient side. To provide evidence of physicians' incentive, we show that the (positive) association is smaller for individuals with good health risk because of the stronger offsetting effect. A detailed discussion is found in Section 5. We note that the empirical evidence on ex ante moral hazard (a reduction of prevention) is very limited, and the consumer incentives are largely reflected as ex post moral hazard (the increase in the utilization of health care under insurance coverage). See Zweifel and Manning (2000) for a comprehensive review of this topic. However, the evidence on the incentives of physicians and interaction with those of consumers has been scanty. Our study also contributes to the literature in this perspective.

The rest of the article is structured as follows: Section 2 discusses the central hypothesis that the empirical analysis is built on. Section 3 proposes the joint modeling framework for discrete outcomes using copulas. Section 4 describes data sources and characteristics of care utilization and insurance choice, as well as explanatory variables. Section 5 summarizes the results. Section 6 concludes. The details of the game-theoretical model and the usage of MEPS data are supplied in the Appendix.

2. HYPOTHESIS

Our central premise is that the health care system encounters the unique challenge of a *two-sided* moral hazard; that is, both patients and providers may have incentives to induce excessive utilization. Traditionally, fee-for-service plans address moral hazard through patient-side cost sharing. By contrast, managed care plans typically employ less patient cost sharing but combine provider-side mechanisms to control moral hazard. On the one hand, compared with traditional plans, managed care gives patients more incentives to overutilize care services, implying a positive association between care utilization and plan type. On the other hand, emphasizing more of provider cost sharing, managed care provides physicians more incentives to limit caregiving, which suggests a negative association between care utilization and plan type. Presumably the incentives from patients and physicians offset each other to some extent. Thus, depending on their relative size, managed care could be associated with either higher or lower care utilization.

The hypothesis of these offsetting incentives is the economic foundation for our subsequent empirical analysis. Though it appears to be intuitive, the existence of two possibilities is not trivial, because patients and physicians will incorporate the action from their counter party into the decision making. For example, expecting reduced care provided through a managed care plan, a patient could pretend to be more sick than he or she actually is to get more care. Thus, we believe that a thorough investigation of health care utilization must study patient and provider behaviors in an integrated manner. To that end, we first develop an economic model that takes a game-theoretical approach, explicitly acknowledging that patients and providers *jointly* determine care utilization under a health care payment system. We show that in equilibrium, managed care may either increase or decrease care utilization. We refer the readers to the Appendix for a detailed discussion of the game-theoretical model.

3. MODELING

3.1. Copula Regression

When examining the effect of a managed care health plan on health care utilization, one potential research issue is the possible endogeneity of the insurance choice in the regression model for care utilization. Such an endogeneity could be introduced by the simultaneity of $utilization$ and $insurance$. The utilization of health care and the selection of health plan are typically jointly determined. On the one hand, the selection of a certain health insurance affects the future care utilization. On the other hand, the expectation of future utilization also determines the health plan choice.

The interest of empirical analysis is to assess to what extent that care utilization and health insurance are interrelated. If the selection of a health plan and utilization of care services are jointly determined, it is appealing to estimate the two equations simultaneously because of the potential efficiency gains compared with estimating the two marginal models separately. However, the estimation could be complicated because the joint distribution of care utilization and insurance choice does not assume an analytical form when the outcomes are discrete and nonlinear. Therefore, we adopt a full information methodology, specifying the joint distribution of $utilization$ and $insurance$ through a parametric copula function and estimating the model using likelihood-based methods.

A copula is simply a multivariate joint distribution defined on a d-dimensional cube $[0, 1]^d$ such that every marginal follows uniform distribution on interval [0, 1]. A copula captures both a linear and nonlinear relationship and has been widely employed in multivariate analysis. The advantage is that it separates the modeling of marginal and dependence structure; see Joe (1997) and Nelsen (2006) for more details on copulas and dependence modeling. In the actuarial science literature, copulas have been found in a variety of applications, such as prediction of nursing home utilization (Sun et al. 2008), modeling insurance claims (Frees and Valdez 2008; Frees et al. 2009), and examination of insurance company expenses (Shi and Frees 2010; Shi 2012). Despite its undeniable popularity, applications of copulas for discrete responses are rarely found. Shi and Valdez (2011) is one example. We also note the extensive applications of copulas in other disciplines, for example, recent work by Smith (2003) and Zimmer and Trivedi (2006) in health economics, Shi et al. (2012) in insurance, and Song et al. (2009) in biostatistics.

Using a copula $H(\cdot)$, the joint distribution of $utilization$ and $insurance$ can be expressed as

$$F(\mathit{utilization}, \mathit{insurance}|\mathbf{x}) = H(F_1(\mathit{utilization}|\mathbf{x}), F_2(\mathit{insurance}|\mathbf{x})|\mathbf{x}), \tag{1}$$

where F and F_j $(j = 1, 2)$ denote the joint and marginal cumulative distribution functions, respectively. Note that the covariate set $\mathbf{x}$ could be different for each marginal. This distinction will be made explicitly in Section 3.2 to indicate the exclusion restriction.

We note that a full information strategy could also be exercised through specifications of a bivariate mixture model. Specifically, the joint distribution of $utilization$ and $insurance$ could be formulated by mixing the conditional independent marginal distributions:

$$f(\mathit{utilization}, \mathit{insurance}|\mathbf{x}) = \int f_1(\mathit{utilization}|\mathbf{x}, \eta_1) f_2(\mathit{insurance}|\mathbf{x}, \eta_2) g(\boldsymbol{\eta}) d\boldsymbol{\eta}.$$

Here $\boldsymbol{\eta} = (\eta_1, \eta_2)'$ denotes the vector of unobserved latent variables and $g(\boldsymbol{\eta})$ represents the joint density function. One commonly used assumption in the literature is the Gaussian heterogeneity (see Train 2003). The downside of the mixture approach is the associated computational difficulty; that is, the maximization of criterion functions often involves numerical or simulation-based approximation. However, to motivate our specification using copulas, we point out that a mixture model in fact has a copula interpretation, with the copula in (1) implicitly defined by

$$\begin{aligned} F(\mathit{utilization}, \mathit{insurance}|\mathbf{x}) &= \sum\sum f(\mathit{utilization}, \mathit{insurance}|\mathbf{x}) \\ &= \int F_1(\mathit{utilization}|\mathbf{x}, \eta_1) F_2(\mathit{insurance}|\mathbf{x}, \eta_2) g(\boldsymbol{\eta}) d\boldsymbol{\eta} \\ &= \int \Pi(F_1(\mathit{utilization}|\mathbf{x}, \eta_1), F_2(\mathit{insurance}|\mathbf{x}, \eta_2)) g(\boldsymbol{\eta}) d\boldsymbol{\eta}, \end{aligned} \tag{2}$$

where $\Pi(\cdot)$ indicates the product copula.

3.2. Model Specification

In this application an individual's choices of insurance and health care utilization are both measured by a discrete random variable. In general, if this pair of discrete variables (y_{i1}, y_{i2}) for the ith individual follows a joint distribution as in Equation (1), their probability mass function can be written as

$$\begin{aligned} f_i(y_{i1}, y_{i2}) &= \text{Prob}(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}) \\ &= H[F_1(y_{i1}), F_2(y_{i2}); \delta] - H[F_1(y_{i1} - 1), F_2(y_{i2}); \delta] \\ &\quad - H[F_1(y_{i1}), F_2(y_{i2} - 1); \delta] + H[F_1(y_{i1} - 1), F_2(y_{i2} - 1); \delta], \end{aligned} \tag{3}$$

where F_1 and F_2 are the cumulative distribution functions of y_{i1} and y_{i2}, respectively. $H(\cdot; \delta)$ is a parametric copula that joins the two marginals with the association parameter δ capturing the relation between the two responses.

In our application we consider two types of health insurance plans, the traditional fee-for-service and managed care, and two categories of health care services, the preventive care and curative care. Let $y_{i,M}$ indicate the plan choice for individual i, and $y_{i,C}$ and $y_{i,P}$ denote his or her utilization of curative care and preventive care, respectively. Their definitions are summarized as follows:

$$\begin{aligned} y_{i,M} &= \begin{cases} 1 \text{ if choose managed care plan} \\ 0 \text{ if otherwise} \end{cases} \\ y_{i,C} &= \text{number of visits to providers for curative care} \\ y_{i,P} &= \begin{cases} 1 \text{ if at least one visits for preventive care} \\ 0 \text{ if no visit} \end{cases}. \end{aligned}$$

Because of the parametric feature of the copula model, one needs specifications of the distribution functions F_1 and F_2 for model inference purposes. As for the distribution regarding health plan choice variable $y_{i,M}$, one could follow the conventional logit or probit formulation:

$$\text{logit}(\pi_i) = \mathbf{x}'_{i,M}\boldsymbol{\beta}_M \text{or probit}(\pi_i) = \mathbf{x}'_{i,M}\boldsymbol{\beta}_M. \tag{4}$$

Here π_i describes the probability of choosing a managed care plan; that is, $\pi_i = \text{Prob}(Y_{i,M} = 1|\mathbf{x}_{i,M})$. Vector $\mathbf{x}_{i,M}$ denotes the vector of covariates that explains the individual's insurance choice, and the coefficient vector $\boldsymbol{\beta}_M$ contains corresponding parameters to be estimated.

We consider standard count distributions by assuming a negative binomial regression model for the number of visits $y_{i,C}$. More specifically, its probability mass function is expressed as

$$f_i(y_{i,C}) = \text{Prob}(Y_{i,C} = y_{i,C}) = \frac{\Gamma(y_{i,C} + \phi)}{\Gamma(\phi)\Gamma(y_{i,C} + 1)} \left(\frac{\phi}{\lambda_i + \phi}\right)^{\phi} \left(\frac{\lambda_i}{\lambda_i + \phi}\right)^{y_{i,C}}. \tag{5}$$

To account for observed individual heterogeneity, a regression is performed through a log link function for its conditional mean, which is given by

$$\text{E}(y_{i,C}|\mathbf{x}_{i,C}) = \lambda_i = \omega_i \exp(\mathbf{x}'_{i,C}\boldsymbol{\beta}_C),$$

where $\mathbf{x}_{i,C}$ denotes the vector of covariates that explains the individual's health care utilization, and the coefficient vector $\boldsymbol{\beta}_C$ contains corresponding parameters to be estimated. The weight parameter ω_i for individual i has the interpretation of the length of time during which the patient is covered by the health insurance for the calendar year in consideration. The dispersion parameter ϕ in the conditional variance expressed as

$$\text{Var}(y_{i,C}|\mathbf{x}_{i,C}) = \lambda_i + \lambda_i^2/\phi$$

provides additional flexibility to fit the count data, especially being able to accommodate either overdispersion or underdispersion often encountered in empirical analysis. See, for example, Cameron and Trivedi (1998).

The utilization of preventive care is also measured by a binary variable; thus, similar to the insurance choice, one could employ either a logit or a probit formation to specify its conditional mean as

$$\text{Prob}(Y_{i,P} = 1|\mathbf{x}_{i,P}) = \frac{1}{1 + \exp(-\mathbf{x}_{i,P}^{'}\boldsymbol{\beta}_P)} \text{ or Prob } (Y_{i,P} = 1|\mathbf{x}_{i,P}) = \Phi(\mathbf{x}_{i,P}^{'}\boldsymbol{\beta}_P), \quad (6)$$

where Φ is the cdf of a standard normal random variable, and vectors $\mathbf{x}_{i,P}$ and $\boldsymbol{\beta}_P$ represent explanatory variables and regression coefficients, respectively.

With the above specification for the bivariate model, the inference could be easily performed using likelihood-based methods. The log likelihood function could be calculated by summing up the logarithm of the joint probability mass function expressed in (3) for all individuals with the set of observable data given by $(y_{i,C}, y_{i,M}, \mathbf{x}_{i,C}, \mathbf{x}_{i,M})$ or $(y_{i,P}, y_{i,M}, \mathbf{x}_{i,P}, \mathbf{x}_{i,M})$. In the log-likelihood function for curative care, $F_C(\cdot)$ and $F_M(\cdot)$ are derived based on negative binomial and logit/probit distributions, respectively. In the log-likelihood function for preventive care, each of $F_P(\cdot)$ and $F_M(\cdot)$ can be either a logit or a probit distribution. Though it is not the focus of this study, it is worth pointing out that the copula approach provides a more flexible framework for the formulation of binary choice models. For example, the bivariate probit model is nested by the copula approach with $F_P(\cdot)$ and $F_M(\cdot)$ following probit and $H(\cdot;\delta)$ being a Gaussian copula. Recall that our interest is the relationship between an individual's health insurance choice and health care utilization. Such a relation is measured by the association parameter δ in the copula function.

One important component in the proposed method is the specification of the copula function, where the association parameter informs the relation between an individual's care utilization and health insurance choice. As explained in Section 2, the two-sided moral hazard, of physicians and patients, could lead to either a positive or a negative association. Without a priori knowledge regarding the direction of the association, we consider the Frank copula, which permits both possibilities:

$$H(u_1, u_2; \delta) = -\frac{1}{\delta}\log\left[1 + \frac{(e^{-\delta u_1} - 1)(e^{-\delta u_2} - 1)}{e^{-\delta} - 1}\right], \quad (7)$$

where δ is the dependence parameter that captures the association between the two responses. The flexibility of allowing for either direction of association has been one of the primary reasons for its popularity in applications in insurance, finance, and biostatistics. It is rather straightforward to show that when $\delta \to 0$, the Frank copula reduces to the product copula, indicating the case of independence. Furthermore, the case with $\delta > 0$ indicates a positive association between the two responses; the reverse is true when $\delta < 0$. In addition, the Frank copula attains both upper and lower Fréchet-Hoeffding bounds. These relationships allow us to understand the type of association that we (empirically) derive from our data. Additional statistical properties of the Frank's family of copulas in (7) have been explored in Genest (1987). We refer readers to this very interesting article to further understand this family of copulas.

In addition to the Frank copula, we further consider three parametric copulas that have been widely used in applied studies, the Farlie-Gumbel-Morgenstern (FGM) copula, the Gumbel copula, and the Clayton copula, as part of the robustness analysis. The cumulative distribution functions of these copulas can be written, respectively, as

$$\begin{aligned}
\text{FGM: } & H(u_1, u_2; \delta) = u_1 u_2 (1 + \delta(1 - u_1)(1 - u_2)),\ |\delta| \leq 1 \\
\text{Gumbel: } & H(u_1, u_2; \delta) = \exp\left[-((-\log u_1)^{\delta} + (-\log u_2)^{\delta})^{1/\delta}\right],\ \delta \geq 1, \\
\text{Clayton: } & H(u_1, u_2; \delta) = \left(u_1^{-\delta} + u_2^{-\delta} - 1\right)^{-1/\delta},\ \delta > 0
\end{aligned}$$

Like the Frank copula, all these copulas nest the product copulas as either a special or a limiting case. However, none of them provides the flexibility of the Frank copula in the dependence modeling. Specifically, both Gumbel and Clayton copulas allow only for positive association. Though the FGM copula permits negative association, it obtains neither Fréchet-Hoeffding bound due to the bounded dependence parameter.

4. DATA

To examine the relation between managed care and health care utilization, this study considers a survey data from the Medical Expenditure Panel Survey (MEPS). Conducted by the Agency for Healthcare Research and Quality (AHRQ), a unit of the Department of Health and Human Services, the MEPS is a set of large-scale surveys of families and individuals, their medical providers, and employers across the United States. It currently consists of two major components; the household component providing data for a probability sample of families and individuals, and the insurance component collecting data from a sample of employers on employer-based health insurance plans.

We use 2008 data from the household component of the MEPS. The data contain a nationally representative sample of the U.S. civilian noninstitutionalized population. Data are collected through an overlapping panel design where measures are taken over two years in five rounds of interviews for each panel drawn from respondents to the National Health Interview Survey. Thus we are looking at essentially a cross-sectional data set though the MEPS is initially designed to be a long panel survey. The data contain detailed information on demographic characteristics, health conditions, health insurance coverage, employment, and most important, a variety of measures of use of medical care services and associated expenditures. The 2008 sample consists of 33,066 individuals. Our study considers a subsample of nonelderly adults (age between 18 and 64) and focuses on individuals who have some form of private insurance and no form of public health insurance, such as Medicaid or Medicare enrollees. This is because we wish to examine the effect of managed care on health care utilization among those who make such choices in private insurance market. Our final sample includes 9,737 individuals.

Our empirical study considers a variety of variables of health care utilization, including three curative care measures and three preventive care measures. Curative care utilization is measured by the number of visits to a certain type of medical providers. In particular, we cover the frequency of visits to doctors including primary care physicians and specialists, to nondoctors such as psychologists, nurse practitioners, and social workers, and to emergency rooms (ERs). The set of preventive care services is measured by binary variables, indicating whether blood pressure was tested in the last year, whether blood cholesterol was checked in the last year, and whether a flu shot was received in the last year.

Managed care is captured via a binary variable. In MEPS each individual is supposed to provide the type of insurance coverage. An individual is assumed to enroll in a managed care plan if he or she reports belonging to a private HMO, or a private gatekeeper plan, or a private plan that has a book or list of doctors. In our sample, about 71% individuals enroll in managed care (MC) plans; the rest 29% belong to nonmanaged care, presumably traditional fee-for-service plans, denoted by FFS in the following presentation.

Table 1 displays the average utilization of both curative care and preventive care services for all enrollees as well as for enrollees in each type of plan. Relative to individuals in traditional health plans, those in managed care have higher doctor and nondoctor visits. The frequency of visits to ERs is comparable for the two plan types. Since the preventive care variables are binary, we interpret the sample mean as the likelihood of using a certain type of care. We observe that persons in managed care plans are more likely to test blood pressure, to have cholesterol checked, and to receive a flu shot.

To provide a more detailed distribution of the number of visits to medical providers, we summarize the frequency of visits to doctors, nondoctors, and ERs in Table 2. As anticipated, the visits to ERs are much less intensive than those to doctors and nondoctors. In fact, most individuals (about 90%) do not visit ERs over the year, while the majority of individuals (about 68%) visit a doctor at least once, and more than 40% of individuals visit a nondoctor at least once. When comparing the utilization under different plan types, one consistently finds that individuals in managed care plans have a higher chance to use medical care, at least for doctor and nondoctor services, than those enrolled in traditional healthplans.

The explanatory variables for the health care utilization and insurance choice equations are selected mainly following Dowd et al. (1991), Goldman et al. (1995), and Deb et al. (2006). These covariates are grouped into three categories: socioeconomic and demographic characteristics, health status, and employment characteristics. Socioeconomic characteristics include age, gender, marital status, ethnicity, education, family size, annual personal income, and location of residence. An individual's health status is captured by a binary variable indicating if the person has a functional limitation, the number of chronic conditions, and a self-perceived health status, which is indicated by five scales representing excellent, very good, good, fair, and poor. Note that some researchers have raised the concern of endogeneity of self-perceived health status in modeling the utilization of medical care; see, for example, Windmeijer and Santos Silva (1997) and Van Ourti (2004). Despite such concern, we follow mainstream studies

TABLE 1
Average Health Care Utilization

	Overall	MC	FFS
Doctor	3.01	3.17	2.62
Nondoctor	1.96	2.09	1.62
ER	0.14	0.14	0.14
Blood pressure	0.79	0.82	0.74
Blood cholesterol	0.54	0.57	0.47
Flu shot	0.29	0.31	0.25
Sample percentage	100	71.30	28.70

TABLE 2
Frequency of Visits to Medical Providers

	Doctor			Nondoctor			ER		
Count	Overall	MC	FFS	Overall	MC	FFS	Overall	MC	FFS
0	3,107	2,018	1,089	5,840	4,034	1,806	8,697	6,196	2,501
1	1,891	1,334	557	1,585	1,160	425	826	601	225
2	1,258	962	296	725	551	174	151	100	51
3	867	641	226	376	274	102	44	32	12
4	626	481	145	242	182	60	9	7	2
5	412	321	91	137	108	29	4	4	0
6	296	227	69	100	78	22	1	0	1
7	246	178	68	85	69	16	3	2	1
8	172	136	36	90	72	18	0	0	0
9	158	116	42	57	42	15	1	0	1
10	125	98	27	42	28	14	1	0	1
11	92	65	27	36	25	11			
12	71	57	14	43	32	11			
13	80	63	17	44	28	16			
14	46	30	16	22	20	2			
15+	290	215	75	313	239	74			
Total	9,737	6,942	2,795	9,737	6,942	2,795	9,737	6,942	2,795

and treat self-perceived health status as an exogenous covariate. To further justify this assumption, we use this health status measure reported in the first found of the survey; thus they are predetermined in our study. The socioeconomic characteristics and health status are included in both equations that determine health care utilization and insurance choice. Employment characteristics, including whether an individual is employed or self-employed, will show only up in the insurance choice equation. In theory, the proposed bivariate copula model can be identified through functional forms; thus no exclusion restrictions are required for identification purposes in the sense of simultaneous equations. However, in the sense of seemingly unrelated regression model, efficiency could be gained when the explanatory variables in different equations are not identical (see Zimmer and Trivedi 2006). Thus we include additional employment characteristics in the insurance choice equation.

The description and summary statistics of covariates are described in Table 3 for the overall sample and by insurance plan. An average person in our sample is an individual of age 41 who has 14 years of education and has a family of three members. Over half of individuals are female, and 63% of them are married. Survey respondents are well diversified in their ethnicity, with 18% Hispanic, 16% black, and 8% Asian. The population are widely distributed over the country, and among them, about 87% live in a metropolitan statistical area, and more than one-third are from the southern United States. Regarding the health status, 15% are found to have certain physical limitation, and most individuals perceive themselves in a healthy state. It is not surprising that we observe a high percentage of employment since our sample is limited to individuals with only private insurance, which is often provided through employers as an employment benefit. When comparing by insurance plans, we notice different socioeconomic characteristics between the enrollees in managed care and traditional plans. Employment characteristics differ too. However, there are fewer differences in observed health status measures across insurance plan types.

5. RESULTS

5.1. Estimation and Inference

This section presents the estimation results for health care utilization and insurance choice equations. The empirical analysis considers three measures of utilization of curative care and three measures of utilization of preventive care. The Frank copula model is estimated for each measure with two marginals: the health care utilization and the choice of health plan. As expected, regression coefficients of insurance choice are robust with respect to different measures of utilization, since the choice of plan types is the same and all models use the same set of explanatory variables. As a result, we choose to present estimates from one of the six models: the joint model of visits to doctors and insurance plan.

TABLE 3
Sample Mean of Explanatory Variables

	Description	Overall	MC	FFS
	Demographic Characteristics			
age	Age in years	41.04	41.62	39.60
female	=1 if person is female	0.52	0.53	0.50
married	=1 if person if married	0.63	0.66	0.56
hispanic	=1 if person is Hispanic	0.18	0.18	0.19
black	=1 if person is black	0.16	0.15	0.17
asian	=1 if person is Asian	0.08	0.08	0.08
edu	Number of years of education	13.63	13.76	13.33
familysize	Family size	3.07	3.07	3.08
income	Annual income	79.75	83.08	71.47
msa	=1 if metropolitan statistical area	0.87	0.89	0.81
northeast	=1 if residence is in Northeast	0.16	0.16	0.14
midwest	=1 if residence is in Midwest	0.22	0.21	0.24
south	=1 if residence is in South	0.36	0.34	0.41
	Health Status			
limitation	=1 if physical limitation	0.15	0.16	0.13
chronic	Number of chronic conditions	1.34	1.38	1.24
excellent	=1 if health is excellent	0.31	0.31	0.32
verygood	=1 if health is very good	0.35	0.36	0.33
good	=1 if health is good	0.25	0.25	0.26
fair	=1 if health is fair	0.07	0.07	0.08
	Employment Characteristics			
employed	=1 if employed	0.85	0.87	0.82
selfemployed	=1 if self-employed	0.07	0.06	0.09

Marginal effects of this model are exhibited in Table 4. We find that older individuals are more likely to choose managed care plans, though the effect is small. Female and married are more likely to enroll in managed care plans than traditional health plans. Black and Asian are found to be less likely in a managed care plan. Individuals who are better educated and who live in metropolitan areas are less likely to select a traditional fee-for-service plan. There are substantive regional differences as well. Interestingly, we observe that self-assessed health status affects insurance choice. In general, healthier individuals tend to choose managed care plans. Such an observation suggests that there might be favorable selection on the basis of unobserved health. It is important to see that both employment-related variables are significant because they are excluded from the utilization equation as exclusion restrictions. It appears that managed care plan enrollment is positively related to being employed and negatively related to self-employment, most likely because of the types of plans available to the self-employed.

The estimation results for the utilization of curative care and preventive care are summarized in Table 5 and Table 6, respectively. The parameter estimates are in general of plausible sign and significance. For curative care utilization, the number of visits is modeled using a negative binomial regression; thus we interpret the regression coefficients as the partial effect of the explanatory variable on the average number of visits. In terms of signs, the effect of most socioeconomic characteristics on curative care utilization is the same for all three measures, though there is some divergence for the utilization of ERs. Overall, age is positively associated with the use of doctor and nondoctor services, and negatively related with the use of ERs. On average, blacks have fewer doctor and nondoctor visits but more ER visits than their nonblack counterparts. Individuals who are better educated and who have a higher income tend to use more doctor and nondoctor services and fewer ER services. In contrast, the effect of health status is quite consistent for the three types of medical care utilization. For both objective and self-perceived health variables, we find that a healthier person is expected to have a lower number of visits to the three types of medical providers.

For preventive care utilization, we examine whether a certain type of care is used over the year via a logit regression model. In this case the regression coefficients could be interpreted as the likelihood of care utilization. Similar to curative care utilization, many socioeconomic characteristics show a significant effect on the use of medical services. However, the directions of these effects are not agreeable when compared across service types. The regional difference is less pronounced in the utilization of

TABLE 4
Estimation Results for Health Plan Choice

	Estimate	SE	p Value
intercept	−0.924	0.021	< 0.001
age	0.006	0.001	< 0.001
female	0.153	0.011	< 0.001
married	0.363	0.018	< 0.001
hispanic	−0.074	0.054	0.171
black	−0.020	0.012	0.097
asian	−0.170	0.085	0.045
edu	0.039	0.004	< 0.001
familysize	−0.025	0.027	0.363
income	2.228	0.008	< 0.001
msa	0.530	0.021	< 0.001
northeast	−0.221	0.021	< 0.001
midwest	−0.389	0.060	< 0.001
southern	−0.443	0.052	< 0.001
limitation	0.166	0.071	0.020
chronic	0.031	0.020	0.125
excellent	0.152	0.006	< 0.001
verygood	0.250	0.010	< 0.001
good	0.112	0.041	0.006
fair	0.077	0.088	0.380
employed	0.416	0.018	< 0.001
selfemployed	−0.659	0.003	< 0.001

preventive care than curative care. In health-related variables, chronic conditions are found to be positively correlated to the likelihood of using all three types of services. However, the effect of self-assessed health status is significant only for the utilization of flu shot, but not for blood pressure and cholesterol.

The main goal of this study is to examine the relationship between managed care and medical care utilization. This relation is reflected by the dependence parameter δ in the Frank copula, which is reported at the bottom of Tables 5 and 6 together with corresponding estimation error. The estimated dependence parameter is positive for all measures of health care utilization, suggesting that a managed care plan is positively associated with the utilization of curative care and preventive care services. For example, δ is roughly 0.53 in the model for doctor visits and 0.60 in the model for blood pressure check. As already alluded to in the previous section, the Frank copula offers the flexibility in accommodating both positive and negative associations. This advantage also makes the hypothesis test rather straightforward because the statistical test for the dependence parameter avoids the problem often encountered with boundary hypothesis. The first test that we perform is a Wald test, with an associated p value reported along with the parameter estimate. A small p-value implies that the dependence parameter is statistically significantly different from zero, indicating a significant positive association between the two responses. We find that except for visits to ERs, the positive relation between managed care plan and medical care utilization is significant for all measures of health care. In addition, one could perform a likelihood ratio test to examine the quality of fit of the Frank copula. Under the standard setup, the chi-square statistic follows a distribution that in this case has one degree of freedom. The observed values of the chi-square statistic and the corresponding p value are also presented at the bottom of both tables. In agreement with the Wald statistic, we find that a significant positive relationship between managed care plan and health care utilization can be observed purging off the effects of exogenous variables.

The positive relationship suggests that individuals in managed care plans tend to use more medical care services compared with their counterparts in traditional fee-for-service plans. This relation could be caused by some unobservable factors that are not accounted for by covariates. As already discussed in Section 2, such latent variables reflect underlying incentives from both enrollees and providers. On the one hand, an enrollee of a managed care plan has more incentive to visit medical providers because typically a managed care plan requires less cost sharing from an enrollee. Such incentive could lead to a positive association. On the other hand, because of gatekeeping, a physician in managed care plans is subject to more restriction in soliciting visits

TABLE 5
Estimation Results for Curative Care Utilization

	Doctor			Nondoctor			ER		
	Estimate	SE	p Value	Estimate	SE	p Value	estimate	SE	p Value
intercept	−0.196	0.006	< 0.001	−1.191	0.006	< 0.001	−0.580	0.011	< 0.001
age	0.004	0.001	0.003	0.007	0.001	< 0.001	−0.019	0.003	< 0.001
female	0.622	0.011	< 0.001	0.506	0.008	< 0.001	0.138	0.078	0.077
married	0.271	0.003	< 0.001	0.118	0.001	< 0.001	−0.035	0.011	0.001
hispanic	−0.171	0.036	< 0.001	−0.704	0.130	< 0.001	−0.198	0.095	0.036
black	−0.300	0.008	< 0.001	−0.808	0.120	< 0.001	0.232	0.012	< 0.001
asian	−0.327	0.047	< 0.001	−0.689	0.070	< 0.001	−0.658	0.161	< 0.001
edu	0.060	0.003	< 0.001	0.113	0.001	< 0.001	−0.023	0.012	0.043
familysize	−0.060	0.003	< 0.001	−0.124	0.001	< 0.001	−0.017	0.021	0.417
income	0.389	0.017	< 0.001	1.162	0.001	< 0.001	−1.230	0.027	< 0.001
msa	0.215	0.002	< 0.001	0.282	0.003	< 0.001	0.085	0.016	< 0.001
northeast	0.210	0.016	< 0.001	0.155	0.003	< 0.001	0.261	0.092	0.004
midwest	0.038	0.100	0.702	0.244	0.034	< 0.001	0.189	0.012	< 0.001
southern	0.106	0.010	< 0.001	−0.264	0.003	< 0.001	−0.055	0.073	0.445
limitation	0.404	0.036	< 0.001	0.538	0.057	< 0.001	0.607	0.081	< 0.001
chronic	0.185	0.010	< 0.001	0.195	0.001	< 0.001	0.134	0.016	< 0.001
excellent	−0.894	0.019	< 0.001	−0.860	0.002	< 0.001	−0.967	0.037	< 0.001
verygood	−0.694	0.004	< 0.001	−0.634	0.174	< 0.001	−0.740	0.090	< 0.001
good	−0.617	0.035	< 0.001	−0.541	0.002	< 0.001	−0.584	0.093	< 0.001
fair	−0.325	0.051	< 0.001	−0.346	0.064	< 0.001	−0.201	0.121	0.098
ϕ	0.905	0.011	< 0.001	0.280	0.001	< 0.001	0.406	0.031	< 0.001
δ	0.525	0.017	< 0.001	0.299	0.005	< 0.001	0.096	0.092	0.297
χ^2	37.420	< 0.001	5.860	0.015	0.402	0.526			

from patients compared with traditional fee-for-service plans where physicians have a higher monetary incentive to provide more medical services. Thus a negative association is expected under this case. As a result, the observed positive relation between managed care and health care utilization should be interpreted as a combined result of the moral hazard from both enrollees and physicians, with a dominating effect of enrollee incentives. Further analysis is needed to show the evidence of physician incentives. We will discuss this in more detail in Section 5.4.

5.2. Treatment Effect

This section quantifies the difference in the utilization of health care between managed care and traditional plan using the treatment effect. We illustrate the calculation for an individual with average characteristics, though the method discussed here is applicable to any arbitrarily selected individual.

The dependence parameter in the copula captures the relation between an individual's health care utilization and the type of health plan. The significant positive dependency suggests more visits to medical providers associated with managed care. In particular, the direct effect of managed care plan on medical care utilization could be extracted from the bivariate copula model, though health plan is not an explanatory variable in the utilization equation. More specifically, the distribution of health care utilization conditional on health plan could be derived using Bayes's rule. For example, one has the following relation for curative care variables:

$$f_i(y_{i,C}|y_{i,M}, \mathbf{x}_{i,C}, \mathbf{x}_{i,M}) = \frac{f(y_{i,C}, y_{i,M}|\mathbf{x}_{i,C}, \mathbf{x}_{i,M})}{f(y_{i,M}|\mathbf{x}_{i,M})}. \tag{8}$$

In Equation (8), $f(y_{i,C}, y_{i,M}|\mathbf{x}_{i,C}, \mathbf{x}_{i,M})$ and $f(y_{i,M}|\mathbf{x}_{i,M})$ are determined by Equations (3) and (5), respectively. One notes that the conditional distribution is specific for each individual because of the heterogeneity. To quantify the effect of managed care plan

TABLE 6
Estimation Results for Preventive Care Utilization

	Blood Pressure			Blood Cholesterol			Flu Shot		
	Estimate	SE	p Value	Estimate	SE	p Value	Estimate	SE	p Value
intercept	−0.988	0.382	0.010	−4.158	0.288	< 0.001	−3.751	0.282	< 0.001
age	0.008	0.003	0.003	0.043	0.002	< 0.001	0.025	0.002	< 0.001
female	0.964	0.055	< 0.001	0.340	0.046	< 0.001	0.442	0.048	< 0.001
married	0.327	0.064	< 0.001	0.292	0.055	< 0.001	0.127	0.059	0.030
hispanic	−0.016	0.076	0.831	0.507	0.067	< 0.001	0.022	0.071	0.758
black	0.074	0.082	0.367	0.527	0.070	< 0.001	−0.294	0.074	< 0.001
asian	−0.322	0.098	0.001	0.355	0.087	< 0.001	0.237	0.088	0.007
edu	0.066	0.012	< 0.001	0.071	0.010	< 0.001	0.061	0.010	< 0.001
familysize	−0.090	0.020	< 0.001	−0.026	0.017	0.138	−0.044	0.019	0.019
income	2.701	0.562	< 0.001	1.978	0.460	< 0.001	2.695	0.446	< 0.001
msa	0.097	0.081	0.230	0.191	0.070	0.006	−0.177	0.071	0.013
northeast	0.249	0.090	0.005	0.389	0.074	< 0.001	−0.009	0.075	0.907
midwest	0.100	0.081	0.218	0.011	0.069	0.877	0.049	0.071	0.486
southern	0.035	0.072	0.629	0.163	0.062	0.009	−0.105	0.065	0.104
limitation	0.114	0.095	0.230	−0.112	0.072	0.121	0.087	0.068	0.200
chronic	0.500	0.030	< 0.001	0.402	0.021	< 0.001	0.210	0.018	< 0.001
excellent	−0.171	0.325	0.599	0.131	0.222	0.554	0.466	0.208	0.025
verygood	−0.045	0.323	0.890	0.227	0.219	0.300	0.441	0.204	0.031
good	−0.010	0.323	0.975	0.267	0.219	0.222	0.357	0.202	0.078
fair	0.119	0.338	0.724	0.247	0.229	0.281	0.416	0.211	0.049
δ	0.602	0.119	< 0.001	0.504	0.104	< 0.001	0.294	0.109	0.007
χ^2	25.864	< 0.001	23.790	< 0.001	7.392	0.007			

on health care utilization, we demonstrate with the average treatment effect (ATE) of being enrolled in different plans as

$$\text{ATE}_C = \text{E}[y_C|y_M = 1, \bar{\mathbf{x}}_{i,C}, \bar{\mathbf{x}}_{i,M}] - \text{E}[y_C|y_M = 0, \bar{\mathbf{x}}_{i,C}, \bar{\mathbf{x}}_{i,M}], \tag{9}$$

where $\bar{\mathbf{x}}_{i,C}$ and $\bar{\mathbf{x}}_{i,M}$ are set to the values of sample average. Similarly, the average treatment effect on preventive care variables could be shown as

$$\text{ATE}_P = \text{Prob}[y_P|y_M = 1, \bar{\mathbf{x}}_{i,P}, \bar{\mathbf{x}}_{i,M}] - \text{Prob}[y_P|y_M = 0, \bar{\mathbf{x}}_{i,P}, \bar{\mathbf{x}}_{i,M}]. \tag{10}$$

The treatment effect could be interpreted as the difference in the expected number of visits to medical providers for curative care utilization, and the difference in the probability of using care for preventive care utilization.

The average treatment effect of managed care plan on curative care utilization is calculated based on (9) for a variety of hypothetical individuals and is reported in Table 7. The hypothetical individuals are assumed to have average characteristics of the entire sample as well as subsamples. In particular, the subsamples include male individuals, female individuals, Hispanic individuals, black individuals, and Asian individuals. The treatment effects are also calculated at the average characteristics of the subsamples of individuals without physical limitation and those with physical limitation, of individuals without chronic conditions, and of those with at least one chronic condition. Finally, we calculate treatment effect at average characteristics of those who actually enrolled in managed care plans and those who actually enrolled in traditional fee-for-service plans.

We observe positive treatment effect of managed care plans for curative care variables, as already implied by the positive association parameter in the corresponding Frank copula model. For example, for those individuals with average characteristics, enrollees in managed care plans are predicted to have 17% more visits to doctors and 12% more to nondoctors. The effect on ER is negligible because the relation between managed care and use of ERs is not statistically significant as shown in the previous section. The size of the treatment effect for the average individual of the entire sample is comparable to the average individual of

TABLE 7
Treatment Effect for Curative Care Utilization

	Doctor			Nondoctor			ER		
	FFS	MC	ATE	FFS	MC	ATE	FFS	MC	ATE
average	2.182	2.561	0.379	1.228	1.385	0.157	0.114	0.119	0.005
male	1.557	1.840	0.282	0.932	1.053	0.121	0.105	0.110	0.005
female	2.970	3.469	0.499	1.581	1.780	0.199	0.122	0.128	0.005
hispanic	1.662	1.961	0.299	0.588	0.666	0.078	0.102	0.106	0.005
black	1.821	2.144	0.323	0.677	0.766	0.089	0.157	0.164	0.007
asian	1.702	2.008	0.306	0.924	1.043	0.120	0.055	0.057	0.003
limitation	4.674	5.430	0.756	3.215	3.611	0.396	0.247	0.257	0.010
nolimitation	1.907	2.244	0.337	1.036	1.170	0.133	0.099	0.104	0.005
Chronic = 0	1.444	1.709	0.265	0.762	0.862	0.100	0.090	0.094	0.004
Chronic > 0	2.856	3.338	0.482	1.678	1.889	0.211	0.132	0.138	0.006
In MC	2.266	2.659	0.392	1.299	1.464	0.165	0.113	0.118	0.005
In FFS	1.986	2.334	0.348	1.069	1.206	0.137	0.116	0.121	0.005

those actually enrolled in managed care plans. For the visits to doctors and nondoctors, the treatment effect is uniformly smaller for average male individuals than average female individuals, for average Hispanic individuals than average non-Hispanic individuals, and for average individuals without physical limitation or chronic conditions than average individuals with a physical limitations or chronic conditions, respectively.

Table 8 summarizes the treatment effect on preventive care utilization according to relation (10). Similar to curative care variables, the effects are quantified for hypothetical individuals with average characteristics from different groups of enrollees. Because the response variables are binary, the treatment effect represents the change in the likelihood of using preventive care services. In general, enrollees in managed care plans have a higher chance of receiving curative care. For example, enrollees with average characteristics in a managed care plan are 4% more likely to receive blood pressure check, 6% more likely to check blood cholesterol, and 3% more likely to get a flu shot. Again, the magnitude of the treatment effect at the average characteristics of the entire sample is very close to those obtained at the average characteristics of those actually enrolled in managed care plans. There is no uniform relation in the treatment effects on the three types of preventive care. For example, for blood pressure and cholesterol checks, the treatment effects are higher for average male individuals than average female individuals, for average Hispanic individuals than average non-Hispanic individuals, and for average individuals without physical limitation or chronic

TABLE 8
Treatment Effect for Preventive Care Utilization

	Blood Pressure			Blood Cholesterol			Flu Shot		
	FFS	MC	ATE	FFS	MC	ATE	FFS	MC	ATE
average	0.809	0.851	0.042	0.513	0.575	0.062	0.249	0.278	0.029
male	0.718	0.774	0.056	0.462	0.525	0.063	0.208	0.234	0.025
female	0.871	0.901	0.030	0.560	0.620	0.061	0.292	0.323	0.031
hispanic	0.745	0.797	0.052	0.495	0.558	0.063	0.207	0.232	0.025
black	0.820	0.860	0.040	0.590	0.649	0.059	0.188	0.212	0.024
asian	0.738	0.792	0.053	0.522	0.584	0.062	0.298	0.329	0.032
limitation	0.909	0.930	0.022	0.661	0.715	0.053	0.342	0.376	0.034
nolimitation	0.785	0.831	0.046	0.486	0.549	0.063	0.235	0.263	0.028
Chronic = 0	0.657	0.721	0.064	0.320	0.377	0.057	0.178	0.201	0.023
Chronic > 0	0.877	0.905	0.028	0.641	0.697	0.055	0.305	0.337	0.032
In MC	0.818	0.858	0.040	0.529	0.591	0.062	0.258	0.287	0.029
In FFS	0.786	0.831	0.046	0.473	0.536	0.063	0.229	0.256	0.027

conditions than average individuals with a physical limitations or chronic conditions, respectively. However, we observe the reverse relation regarding the size of the treatment effect for flu shot utilization.

The result of treatment effects reflects, on the one hand, the heterogeneity in the responses of different groups of enrollees to the incentives of care implied by different models of health care provision. On the other hand, it also exhibits the reaction of physicians to the incentives and restrictions on care provision under different health plans. As discussed in Section 5.1, the parameter δ in the copula captures the effect of managed care on health care utilization. Such a relation combines two separate effects: first, the incentives of enrollees, including the direct causal effect on utilization of being enrolled in different plans and the indirect effect of insurance choice due to the expectation of future utilization. In addition, δ also reflects the incentives of physicians in different insurance plans. In general, the enrollees in managed care plans have higher incentive of over-utilization of medical care services because of the cost-sharing mechanism. In contrast, the physicians in managed care plans are subject to more restrictions and thus have less incentive for excessive care delivery. The dependence parameter δ reflects the net effects of the two offsetting incentives. This also explains the relatively small magnitude of treatment effects.

5.3. Model Validation

One advantage of the copula modeling framework is that the copula preserves the marginals while accommodating the association among them. This property allows us to separately assess the goodness-of-fit for the marginals and the copula. For marginal distributions, we compare observed and fitted frequencies for all response variables. The results for curative care and preventive care variables are presented in Table 9 and Table 10, respectively. Because of the heterogeneity among individuals, the mass probability for the count and binary random variables is different across individuals. We calculate the probability at all possible values in their support for each enrollee and present the average over all enrollees in both tables. For visits to doctors and nondoctors, we combined the probabilities for the number of visits above 15 due to the small number of observations in each category. The consistency between the actual and fitted frequencies implies a very satisfactory fit for each marginal. Though not reported here, similar comparison is also performed for insurance plan choice, and the result suggests a favorable fit as well.

In this study the joint distribution of care utilization and insurance choice is constructed via parametric copulas. Though the model-based approach enjoys the straightforward implementation of estimation and inference using likelihood-based methods, the estimates of the complex copula regression model could be sensitive to sample variations and distributional assumptions. To address such concerns, we perform two types of robustness checks: one for the choice of sample coverage, the other for the specification of copulas. Regarding the joint distribution, our primary concern is whether the relation between two marginals (care utilization and insurance choice) are well captured by the copula function, rather than the specific form of the parametric copula. Hence, the robustness tests focus on the significance of the association and its implications on utilization difference, which is quantified by the treatment effect.

TABLE 9
Goodness-of-Fit for Curative Care Utilization

	Doctor			Nondoctor			ER	
Count	Observed	Fitted	Count	Observed	Fitted	Count	Observed	Fitted
0	0.319	0.321	0	0.600	0.614	0	0.893	0.893
1	0.194	0.193	1	0.163	0.132	1	0.085	0.084
2	0.129	0.126	2	0.075	0.067	2	0.016	0.016
3	0.089	0.087	3	0.039	0.042	3	0.005	0.004
4	0.064	0.061	4	0.025	0.028	4	0.001	0.001
5	0.042	0.044	5	0.014	0.021	5	0.000	0.001
6	0.030	0.033	6	0.010	0.015	6	0.000	0.000
7	0.025	0.025	7	0.009	0.012	7	0.000	0.000
8	0.018	0.019	8	0.009	0.009	8	0.000	0.000
9	0.016	0.015	9	0.006	0.008	9	0.000	0.000
10	0.013	0.012	10	0.004	0.006	10	0.000	0.000
11	0.009	0.010	11	0.004	0.005			
12	0.007	0.008	12	0.004	0.004			
13	0.008	0.006	13	0.005	0.004			
14	0.005	0.005	14	0.002	0.003			
15+	0.030	0.034	15+	0.032	0.029			

TABLE 10
Goodness-of-Fit for Preventive Care Utilization

	Blood Pressure			Blood Cholesterol			Flu Shot	
	Observed	Fitted		Observed	Fitted		Observed	Fitted
0	0.207	0.206	0	0.457	0.458	0	0.709	0.711
1	0.793	0.794	1	0.543	0.542	1	0.291	0.289

The first robustness test involves the influence of sample coverage on model inference. Recall that employment status is one important explanatory variable in the determination of health plan choices. In the health insurance market, most private plans are available to individuals through their employers. Thus, the employment status might be jointly determined with the preference to access to a particular type of plan, at least for those who are employed. As a result, we reestimate the Frank copula model with a subsample of individuals excluding only unemployed and a subsample of individuals excluding both unemployed and self-employed. This exercise is carried out for the three models of curative care utilization as well as the three models of preventive care utilization. The results are displayed in Tables 11 and 12, respectively.

In both tables we present the sample size, the dependence test for the association parameter in the copula, the expected utilization of average individuals, and the treatment effects. The results using the entire sample with a Frank copula are reproduced in both tables for comparison purposes. Consistently, for both subsamples, a significant positive relation between health care utilization and managed care is found except for the visits to ERs. The size and sign of treatment effects are also comparable to the inference using the entire original sample.

Our second robustness check examines the sensitivity of parameter estimates to different copula specifications. In the copula regression model, we have focused on the Frank copula because of its flexibility and comprehensiveness for dependency modeling. In particular, we refit the model to the entire original sample using the FGM, Gumbel, and Clayton copulas. However, as pointed

TABLE 11
Robust Test for Curative Health Care Utilization

	Dependence Test			Treatment Effect		
	N	χ^2 Statistic	p Value	FFS	MC	ATE
Doctor						
Sample without unemployed	8,324	24.484	< 0.001	2.099	2.445	0.347
Sample without unemployed and self-employed	7,618	25.632	< 0.001	2.067	2.437	0.371
Full sample with FGM copula	9,737	36.180	< 0.001	2.186	2.559	0.373
Full sample with Gumbel copula	9,737	21.920	< 0.001	2.242	2.537	0.295
Full sample with Clayton copula	9,737	37.600	< 0.001	2.208	2.551	0.343
Full sample with Frank copula	9,737	37.420	< 0.001	2.182	2.561	0.379
Nondoctor						
Sample without unemployed	8,324	3.995	0.046	1.290	1.400	0.111
Sample without unemployed and self-employed	7,618	3.113	0.078	1.253	1.345	0.092
Full sample with FGM copula	9,737	6.120	0.013	1.244	1.379	0.135
Full sample with Gumbel copula	9,737	4.560	0.033	1.257	1.374	0.116
Full sample with Clayton copula	9,737	2.820	0.093	1.239	1.381	0.142
Full sample with Frank copula	9,737	5.860	0.015	1.228	1.385	0.157
ER						
Sample without unemployed	8,324	0.072	0.788	0.120	0.118	−0.002
Sample without unemployed and self-employed	7,618	0.108	0.742	0.121	0.119	−0.003
Full sample with FGM copula	9,737	0.398	0.528	0.114	0.119	0.005
Full sample with Gumbel copula	9,737	0.050	0.823	0.116	0.118	0.002
Full sample with Clayton copula	9,737	0.324	0.569	0.113	0.119	0.006
Full sample with Frank copula	9,737	0.402	0.526	0.114	0.119	0.005

TABLE 12
Robust Test for Preventive Health Care Utilization

	Dependence test			Treatment effect		
	N	χ^2 Statistic	p Value	FFS	MC	ATE
Blood pressure						
Sample without unemployed	8,324	18.632	< 0.001	0.806	0.845	0.039
Sample without unemployed and self-employed	7,618	17.134	< 0.001	0.808	0.847	0.039
Full sample with FGM copula	9,737	25.834	< 0.001	0.810	0.851	0.041
Full sample with Gumbel copula	9,737	26.124	< 0.001	0.810	0.851	0.041
Full sample with Clayton copula	9,737	21.704	< 0.001	0.808	0.851	0.043
Full sample with Frank copula	9,737	25.864	< 0.001	0.809	0.851	0.042
Blood cholesterol						
Sample without unemployed	8,324	18.762	< 0.001	0.516	0.577	0.060
Sample without unemployed and self-employed	7,618	22.047	< 0.001	0.505	0.574	0.069
Full sample with FGM copula	9,737	24.010	< 0.001	0.513	0.575	0.063
Full sample with Gumbel copula	9,737	24.450	< 0.001	0.516	0.574	0.058
Full sample with Clayton copula	9,737	16.070	< 0.001	0.522	0.572	0.049
Full sample with Frank copula	9,737	23.790	< 0.001	0.513	0.575	0.062
Flu shot						
Sample without unemployed	8,324	7.292	0.007	0.250	0.281	0.031
Sample without unemployed and self-employed	7,618	7.974	0.005	0.252	0.287	0.035
Full sample with FGM copula	9,737	7.392	0.007	0.249	0.278	0.029
Full sample with Gumbel copula	9,737	8.492	0.004	0.249	0.278	0.030
Full sample with Clayton copula	9,737	5.172	0.023	0.254	0.276	0.022
Full sample with Frank copula	9,737	7.392	0.007	0.249	0.278	0.029

out in Section 3.2 these copulas have limitations in modeling association, but they are suitable for the mild positive dependence in our application. The results on dependence test and treatment effect are exhibited in Tables 11 and 12, respectively, for curative care and preventive care variables. Generally the influence of copula specification is fairly small, and the curative care variables show higher sensitivity than preventive care variables.

5.4. Discussion on Moral Hazard

The moral hazard from enrollees implies a positive relation between managed care plans and the utilization of medical care services, while the moral hazard from medical providers implies a negative relationship. Since the incentives of patients and physicians offset each other, the dependence parameter in the copula represents a net effect. Thus, the significant positive association between care utilization and plan type reflects only the incentive from patients, which dominates physician incentives. To further provide evidence of physicians' incentives, we reexamine the relation of types of health plan and care utilization for different groups of individuals. Specifically, we divide the entire sample into two groups according to individuals' health risk: the good risk group and the bad risk group. Our hypothesis is that in the presence of physicians' moral hazard, the association between care utilization and type of health plan is weaker or could even become negative for the good risk group compared to the bad risk group. This is because good risks are healthier individuals for whom physicians have more discretion on care provision than bad risks. One could think of the gatekeeping mechanism in managed care plans, where fewer restrictions could be applied to individuals with poor health. When physicians can limit care providing to a greater extent, the negative relation between care usage and managed care is stronger, and so is the offsetting effect. Depending on the relative magnitude of physicians' incentive, the resulting association for the group of good risks will be smaller, be it either positive or negative.

In our analysis, we group individuals using self-perceived health status. Individuals with excellent, very good, and good scales are classified into good risks, and individuals with fair or poor scales are classified into bad risks. The Frank copula model is recalibrated for the two risk groups. We report the estimated dependence parameter in the Frank copula and the associated independence test in Table 13. For both curative care and preventive care variables, the association between managed care plans and utilization of medical care services is consistently smaller for good risks than bad risks, suggesting the existence of physicians' moral hazard.

TABLE 13
Independence Test for Subsamples

	δ	SE	χ^2 Statistic	p Value
Doctor				
Good risk	0.474	0.054	27.440	< 0.001
Bad risk	1.128	0.296	15.853	< 0.001
Nondoctor				
Good risk	0.151	0.002	3.995	0.046
Bad risk	0.631	0.249	3.953	0.047
Blood pressure				
Good risk	0.546	0.123	19.924	< 0.001
Bad risk	1.502	0.511	9.166	0.002
Blood cholesterol				
Good risk	0.472	0.108	19.169	< 0.001
Bad risk	0.775	0.373	4.378	0.036

5.5. Limitation

We proposed a copula regression model to examine the relation between care utilization and health insurance plan using the MEPS survey data, and we have demonstrated several useful applications of the copula approach. However, the analysis has its own limitations. First, the MEPS uses a complex sampling design to collect data where a sampling weight is assigned to each sampling unit to reflect the survey design. The current study has focused on the copula model and thus excluded the sampling weight. Therefore one needs to be careful when drawing "policy-type" implications from the results. Second, the sample design of the MEPS involves stratification, clustering, multiple stages of selection, and disproportionate sampling. The empirical analysis in this study is completed at the individual-person level. We ignored the potential cluster effect within a household since the primary interest is the person-level care utilization. Finally, because of the data limitation, we performed a special statistical test to support evidence of physician incentives. Ideally one needs episode data to formally separate the moral hazard from patients and medical providers.

In summary, we sampled and analyzed the MEPS data mainly following the existing literature in actuarial sciences and health economics, with an emphasis on demonstrating the the application of the copula model in the empirical analysis. The results in our analysis are limited to the modeling technique and the particular sample of individuals. One should be careful in extrapolating the implications for a larger population and in drawing policy-type conclusions.

6. CONCLUDING REMARKS

The moral hazard of both patients and physicians under a certain health plan affects care utilization. Motivated by such observation, we examined the relation between the type of health plan and individuals' utilization of medical care services. In particular, we compared care utilization between managed care and traditional fee-for-service plans, where incentives from both consumers and medical providers differ. We emphasized the endogeneity issue introduced by the simultaneity of individuals' insurance choice and care usage. Adopting a full information approach, we employed a copula to jointly model the two outcomes conditional on explanatory variables, where the association parameter revealed the effect of managed care on medical care utilization.

To test our hypothesis, we applied the copula approach to a U.S. population sample from the MEPS data. We limited the analysis to individuals with some form of private insurance but no coverage from public plans so that the implications could be applied to the private health insurance market. As for the measures of care utilization, we examined three indicators for preventive care, whether an individual took blood pressure test, blood cholesterol check, and flu shot, and three indicators for curative care: the number of visits to doctors, nondoctors, and ERs. We found that managed care is related to either higher likelihood or higher frequency of using medical care, except for ERs. Such a positive relation is consistent with the moral hazard hypothesis. In fact, it represents a combined effect of incentives from both patients and physicians. Furthermore, we invoked the theory to identify subgroups of policyholders across whom one may expect different degrees of incentives. A series of robustness tests were performed, and the results supported the empirical findings.

It is interesting to find that managed care is associated with higher frequency of visits to medical providers. Presumably managed care is a more cost-efficient design for health care delivery system; thus an intriguing question to ask is "does more frequent visits lead to less health care expenditures?" We leave this as an idea for future research.

REFERENCES

Cameron, A. C., and P. K. Trivedi. 1998. *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.

Deb, P., C. Li, P. Trivedi, and D. Zimmer. 2006. The Effect of Managed Care on Use of Health Care Services: Results from Two Contemporaneous Household Surveys. *Health Economics* 15(7): 743–760.

Deb, P., and P. Trivedi. 2006. Specification and Simulated Likelihood Estimation of a Non-Normal Treatment-Outcome Model with Selection: Application to Health Care Utilization. *Econometrics Journal* 9(2): 307–331.

Dowd, B., R. Feldman, S. Cassou, and M. Finch. 1991. Health Plan Choice and the Utilization of Health Care Services. *Review of Economics and Statistics* 73(1): 85–93.

d'Uva, T. 2006. Latent Class Models for Utilisation of Health Care. *Health Economics* 15(4): 329–343.

Executive Office of the President, Council of Economic Advisors. 2009. *The Economic Case for Health Care Reform*. http://www.whitehouse.gov/assets/documents/CEA_Health_Care_Report.pdf

Frees, E., and E. Valdez. 2008. Hierarchical Insurance Claims Modeling. *Journal of the American Statistical Association* 103 (484): 1457–1469.

Frees, E., J. Gao, and M. Rosenberg. 2011. Predicting the Frequency and Amount of Health Care Expenditures. *North American Actuarial Journal* 15(3): 377–392.

Frees, E., P. Shi, and E. Valdez. 2009. Actuarial Applications of a Hierarchical Insurance Claims Model. *ASTIN Bulletin* 39(1): 165–197.

Frees, E. W. 2010. *Regression Modeling with Actuarial and Financial Applications*. Cambridge: Cambridge University Press.

Genest, C.. 1987. Frank's Family of Bivariate Distributions. *Biometrika* 74(3): 549–555.

Glied, S.. 2000. Managed Care. In *Handbook of Health Economics*, edited by A. Culyer and J. Newhouse, Vol. 1, pp. 707–753. Elsevier.

Goldman, D., S. Hosek, L. Dixon, and E. Sloss. 1995. The Effects of Benefit Design and Managed Care on Health Care Costs. *Journal of Health Economics* 14(4): 401–418.

Jiménez-Martín, S., J. Labeaga, and M. Martínez-Granado. 2002. Latent Class Versus Two-Part Models in the Demand for Physician Services Across the European Union. *Health Economics* 11(4): 301–321.

Joe, H.. 1997. *Multivariate Models and Dependence Concepts*. New York: Chapman & Hall/CRC.

Miller, R., and H. Luft. 1994. Managed care plan performance since 1980. *Journal of the American Medical Association* 271(19): 1512–1519.

Nelsen, R.. 2006. *An Introduction to Copulas*. New York: Springer.

Shi, P.. 2012. Multivariate Longitudinal Modeling of Insurance Company Expenses. *Insurance: Mathematics and Economics* 51(1): 204–215.

Shi, P., and E. Frees. 2010. Long-Tail Longitudinal Modeling of Insurance Company Expenses. *Insurance: Mathematics and Economics* 47(3): 303–314.

Shi, P., and E. Valdez. 2011. A Copula Approach to Test Asymmetric Information with Applications to Predictive Modeling. *Insurance: Mathematics and Economics* 49(2): 226–239.

Shi, P., W. Zhang, and E. Valdez. 2012. Testing Adverse Selection with Two-Dimensional Information: Evidence from the Singapore Auto Insurance Market. *Journal of Risk and Insurance* 79(4): 1077–1114.

Smith, M.. 2003. Modelling Sample Selection Using Archimedean Copulas. *Econometrics Journal* 6(1): 99–123.

Song, P. X.-K., M. Li, and Y. Yuan. 2009. Joint Regression Analysis of Correlated Data Using Gaussian Copulas. *Biometrics* 65(1): 60–68.

Sun, J., E. Frees, and M. Rosenberg. 2008. Heavy-Tailed Longitudinal Data Modeling Using Copulas. *Insurance Mathematics and Economics* 42(2): 817–830.

Train, K.. 2003. *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.

Van Ourti, T. 2004. Measuring Horizontal Inequity in Belgian Health Care Using a Gaussian Random Effects Two Part Count Data Model. *Health Economics* 13(7): 705–724.

Windmeijer, F., and J. Santos Silva. 1997. Endogeneity in Count Data Models: An Application to Demand for Health Care. *Journal of Applied Econometrics* 12(3): 281–294.

Zimmer, D., and P. Trivedi. 2006. Using Trivariate Copulas to Model Sample Selection and Treatment Effects: Application to Family Health Care Demand. *Journal of Business and Economic Statistics* 24(1): 63–76.

Zweifel, P., and W. Manning. 2000. Moral Hazard and Consumer Incentives in Health Care. In *Handbook of Health Economics*, edited by A. Culyer and J. Newhouse, Vol. 1, pp. 409–459. Elsevier.

Discussions on this article can be submitted until July 1, 2014. The authors reserve the right to reply to any discussion. Please see the Instructions for Authors found online at http://www.tandfonline.com/uaaj for submission instructions.

APPENDIX

A.1. Theoretical Model

We develop an economic model of health care utilization to study how care used is influenced by payment arrangements of health insurance. Specifically, we use a game-theoretic approach to study how patients (consumers) and health care providers *jointly* determine the amount of care utilization, given an insurance payment system. An insurance payment system uses a coinsurance rate to delineate patients' incentive to demand care and a reimbursement method to structure providers' incentive to supply care. We present the interaction of patients and providers as a sequential game, in which patients first decide whether to seek care and providers subsequently determine the amount of care supplied if they receive a patient. Throughout this section, male pronouns will be used to indicate patients, and female pronouns physicians.

Individual characteristics will be controlled for in the empirical analysis; for a clean exposition they are suppressed in the economic model. Consider a representative consumer with a state-dependent utility function $U(X, h, s)$, where $X \geq 0$ denotes consumption, $h \geq 0$ denotes health care used, and s refers to the consumer's health state. Let s be a realization of the underlying continuous random variable S with cumulative distribution function $F(\cdot)$ and support [0, 1]. To assist the development of intuition, consider higher s as worse health state. We assume a quasilinear utility form with consumption X being the numeraire good:

$$U(X, h, s) = X + u(h, s). \tag{A.1}$$

The following regularity conditions are imposed on the utility from health care services: $u(\cdot, \cdot)$ is continuous, twice differentiable, and $u(\cdot, s)$ is concave for any $s \in [0, 1]$. $u(\cdot, s)$ is increasing at least for some range of h, but we do not rule out the possibility that too much health care can cause disutility. Because higher s indicates worse health state, it is also natural to assume $\partial^2 u/\partial h \partial s \geq 0$; that is, the marginal utility gain from more health care is higher in a worse health state.

The cost of health care provision is given by function $C(h, s)$. $C(\cdot, \cdot)$ is assumed to be continuous and twice differentiable. $C(\cdot, s)$ is convex for any $s \in [0, 1]$, and $\partial^2 C/\partial h \partial s \geq 0$, that is, the marginal cost of health care is larger for a worse health state. In the first stage of the game, a consumer first makes a binary choice A on whether to initiate the care process. Define an indicator function $\mathbb{I}_A(s) = 1$ if he decides to seek care at health state s, and $\mathbb{I}_A(s) = 0$ otherwise. If the patient decides to go see a provider, he must share partially the cost of the subsequent care. For simplicity, an insurance payment system is assumed to use a coinsurance rate $0 \leq \alpha \leq 1$ to determine patient cost sharing. So the consumer's budget constraint can be written as

$$Y \geq P + X + \alpha C(h, s)\mathbb{I}_A(s), \tag{A.2}$$

where $Y > 0$ is income, $P > 0$ the insurance premium, and the price of the numeraire good X is normalized to one. Assume Y is large enough so that one always consumes a positive amount of X. In other words, there is no income effect on health care demanded.

If a consumer decides not to go see a provider in the first stage of the game, the game effectively ends with zero health care utilization. If he does go to a provider to seek care, the exact amount of care supplied is determined by the provider. Ultimately, a health care provider is concerned about profits. However, for legal and ethical reasons, she must also to some extent care about patient welfare. We construct a provider's objective function as

$$V(h, s) = u(h, s) + R - \beta C(h, s). \tag{A.3}$$

The first term captures the provider's concern about patient welfare, as an agent the patient hires to deliver care services. > 0 measures the degree of agency. The bigger is, the more weight patient welfare carries in the objective function, relative to profits. The reimbursement for care supplied consists of two components: a lump-sum payment $R \geq 0$ and a provider cost-sharing fraction β. Parameter β is negative if a provider earns some profits marginally by treating a patient. So we assume the range of β to be $[\beta_0, 1]$ where β_0 may be negative or zero.

We now turn to the solution of the game described above. To solve this sequential game, we shall invoke the concept of subgame perfect equilibrium. We start by characterizing how patients and providers optimally behave, in each health state s. These can then be used to calculate the expected health care utilization.

By the method of backward induction, a provider's optimal choice is first characterized. Denote by $h^*(s)$ the optimal amount of care supplied, given that she receives a patient with health state s. With objective function (13), $h^*(s)$ is determined implicitly by

$$u_h(h^*(s), s) - \beta C_h(h^*(s), s) = 0 \;\; h^*(s) > 0, \text{ otherwise } h^*(s) = 0. \tag{A.4}$$

Subscripts of a function denote partial derivatives in the above equation. It is straightforward to see the effect of provider cost sharing on the induced health care utilization: $h^*(s)$ decreases as β increases, with concavity of $u(\cdot, s)$ and convexity in $C(\cdot, s)$. That is, increasing provider cost sharing on the margin decreases their incentive to supply care.

In the first stage of the game by (11) and (12), a consumer optimally chooses to go see a provider, that is, $\mathbb{I}_{A^*}(s) = 1$ if

$$u(h^*(s), s) - u(0, s) > \alpha C(h^*(s), s), \tag{A.5}$$

anticipating future care received $h^*(s)$. Otherwise, $\mathbb{I}_{A^*}(s) = 0$. Condition (15) shows that the consumer is less likely to seek care if the coinsurance rate α is higher.

The equilibrium outcome, given a health state s, is completely characterized by the above $h^*(s)$ and $\mathbb{I}_{A^*}(s)$. Finally, we can compute the expected health care utilization by

$$E_S(h) = \int_0^1 \mathbb{I}_{A^*}(s)h^*(s)dF(s). \tag{A.6}$$

To summarize, we have constructed a model that derives the expected health care utilization given an insurance payment system. The focus has been investigating how a payment system affects the incentives of patients and providers, who jointly determine the consumption of health services. The economic model sheds light on the question of whether managed care increases or decreases utilization compared to fee-for-service plans, with a different design of payment system. Typically, managed care plans offer more generous cost sharing rules to patients (e.g., lower α) and incorporate more restrictions on provider reimbursement (e.g., higher β). Using the framework of this model, it is easy to construct examples where managed care may either increase or decrease utilization. In the empirical analysis that ensues, we will put it to the test.

A.2. MEPS Data

For the purposes of replicating the results, Table A.1 provides a detailed description of the MEPS variables and a translation of these variables into the ones used in our analysis.

TABLE A.1
Description and Usage of MEPS Variables

MEPS Variables	Description	Our Variables
OBDRV08, OPDRV08	Number of physician visits	doctor
OBOTHV08, OPOTHV08	Number of nonphysician visits	nondoctor
ERTOT08	ER visits	er
BPCHEK53	Indicator of blood pressure test	blood pressure
CHOLCK53	Indicator of blood cholesteroal test	blood cholesterol
FLUSHT53	Indicator of flu shot	flu shot
PRVHMO31, PHMONP31	Indicators of managed care	managed care
PRVMNC31, PRVMNC31		fee-for-service
PRVDRL31, PRDRNP31		
AGE31X	Age of the person	age
SEX	Gender of the person	female
MARRY31X	Marrital status	married
RACETHNX	Race of the person	hispanic, black, asian
EDUCYR	Number of years of education	edu
FAMSZEYR	Family size	familysize
FAMINC08	Annual income	income
MSA08	Metropolitan statistical area	msa
REGION08	Residence region of the person	northeast, midwest
		south
ANYLIM08	Physical limitation	limitation
HIBPDX, CHDDX, ANGIDX	Indicator of chronic condition	chronic
EMPHDX, CHBRON31, MIDX		
OHRTDX, STRKDX, DIABDX		
ARTHDX, ASTHDX, CHOLDX		
JTPAIN31, CANCERDX		
RTHLTH31	Self-perceived health	excellent, very good
		good, fair
EMPST31	Indicator of employment	employed
SELFCM31	Indicator of self-employment	selfemployed