Modeling Multivariate Distributions Using Copulas: Applications in Marketing

Article *in* Marketing Science · January 2011

DOI: 10.1287/mksc.1090.0491 · Source: RePEc

CITATIONS
72

READS
255

2 authors:

Michael Stanley Smith
University of Melbourne
60 PUBLICATIONS **1,864** CITATIONS
SEE PROFILE

Peter J. Danaher
Monash University (Australia)
104 PUBLICATIONS **3,570** CITATIONS
SEE PROFILE

Modeling Multivariate Distributions Using Copulas: Applications in Marketing

Modeling Multivariate Distributions Using Copulas: Applications in Marketing

Peter J. Danaher

Michael S. Smith

Melbourne Business School, University of Melbourne

February 20, 2009

Forthcoming in Marketing Science

Peter J. Danaher is Coles Myer Professor of Marketing and Retailing and Michael S. Smith is Professor of Econometrics at the Melbourne Business School. Authors are listed alphabetically. Address: Melbourne Business School, University of Melbourne, 200 Leicester Street, Carlton, VIC 3053, Australia. Emails: p.danaher@mbs.edu; mike.smith@mbs.edu . The authors thank comScore for providing the online panel data used for part of this study.

Modeling Multivariate Distributions Using Copulas: Applications in Marketing

Abstract

In this research we introduce a new class of multivariate probability models to the marketing literature. Known as "copula models", they have a number of attractive features. First, they permit the combination of any univariate marginal distributions that need not come from the same distributional family. Second, a particular class of copula models, called "elliptical copula", have the property that they increase in complexity at a much slower rate than existing multivariate probability models as the number of dimensions increase. Third, they are very general, encompassing a number of existing multivariate models, and provide a framework for generating many more. These advantages give copula models a greater potential for use in empirical analysis than existing probability models used in marketing. We exploit and extend recent developments in Bayesian estimation to propose an approach that allows reliable estimation of elliptical copula models in high dimensions. Rather than focusing on a single marketing problem, we demonstrate the versatility and accuracy of copula models with four examples to show the flexibility of the method. In every case, the copula model either handles a situation that could not be modeled previously, or gives improved accuracy compared with prior models.

Keywords: Bayesian Estimation; Discrete Copula; Markov chain Monte Carlo; Gaussian Copula; Media Modeling; Probability Models; Website Page Views

1 Introduction

The rapid growth in information technology in business, combined with the relatively low cost of data storage, has resulted in a corresponding explosion in availability of customer data (Rossi and Allenby 2000). Examples of firms that have made good use of their large customer databases include Harrahs Entertainment, Capital One and Netflix (Davenport and Harris 2007). However, such firms are in the minority, with Davenport and Harris (2007, p.24) estimating that fewer than 10% of businesses worldwide routinely make use of their customer data as part of their firm's strategy. Current methods tend to be limited to analyses of one variable at a time, such as purchase amount, number of purchases per year and time since the last purchase. Because these variables are often highly correlated, more useful information can be gained by looking at multivariate rather than univariate distributions. For example, multivariate distributions enable the estimation of partial correlations, the distribution of functions of component variables (ratios, sums, or other metrics) and conditional distributions used for prediction. A major barrier to implementing such multivariate distributions is that sometimes the constituent univariate marginals do not have the same distributional form. For instance, the total amount purchased might have a log-normal distribution, while the time since last purchase might have an exponential distribution. It is challenging to construct a bivariate distribution with these two specific margins.

There are many situations in marketing where data can be modeled with a well-established univariate distribution. Examples include the negative binomial distribution (NBD) for the total number of purchases of a product within a category (Ehrenberg 1988), the beta-binomial distribution for the number of exposures to a TV advertisement (Rust 1986), a proportional hazards model for the time to product purchase (Jain and Vilcassim 1991) and a multinomial logit (MNL) model for brand purchases within a category (Guadagni and Little 1983). However, new data, particularly from the Internet and customer relationship management (CRM) warehouses, requires the simultaneous analysis of several disparate variables (Fader

and Hardie 2007). Examples include combining the number of visits to a website and the duration of each visit (Danaher et al. 2006), the frequency of direct mail send-outs and purchase amounts (Schweidel, Fader and Bradlow 2008a) and a bivariate timing model for customer acquisition and retention (Schweidel, Fader and Bradlow 2008b). That is, many new marketing problems require the combination of univariate distributions that are not from the same family, even including mixtures of discrete and continuous distributions.

To date, methods for combining pontentially different univariate distributions into a multivariate distribution, known as "copula modeling", have appeared sporadically in the statistics literature over the past few decades (see, for example, Genest and Mackay 1986), although applications have largely been limited to continuous distributions in two dimensions. Recently, copula modeling has become popular in the finance and econometric literatures (Poon, Rockinger and Tawn 2004; Cherubini, Luciano and Vecciato 2004; Hong, Tu and Zhou 2007; Trevidi and Zimmer 2007). Here, copula modeling allows for more accurate modeling of the multivariate distribution of asset returns, including inter-dependencies. This proves important in financial investment because it allows for the construction of more efficient portfolios and improved measurement of risk. But, again, applications in finance are usually restricted to continuous distributions in low dimensions. High dimension discrete distributions pose special challenges in copula modeling, where traditional maximum likelihood estimation proves to be infeasible. However, in a new development in the statistical literature, Pitt, Chan and Kohn (2006) show how to employ the power of Bayesian Markov chain Monte Carlo (MCMC) estimation to extend estimation of copula models to high dimensional situations where the margins can be any combination of discrete or continuous distributions.

The purpose of this paper is to demonstrate how copula models can be constructed and used in a variety of marketing applications. Previous marketing examples where copulas can potentially be used include: modeling inter-visit times across websites, as demonstrated in a bivariate setting by Park and Fader (2004); modeling magazine advertising campaigns

across several magazines (Danaher 1991); and modeling page views across multiple websites (Danaher 2007). In more than two dimensions the models previously used in all these studies become cumbersome or computationally challenging. For instance, Danaher's (1991; 2007) models require approximations for even trivariate distributions. We show that copula models require no such approximation and demonstrate how they are empirically superior to prior multivariate models. We follow and extend the work of Pitt et al. (2006) by developing a new and efficient Bayesian MCMC estimation algorithm that allows estimation of a particular class of copula called "elliptical copula". In doing so, we show how elliptical copulas can be used in high dimensional settings with discrete marginal distributions. The ability to do so now makes the use of copula modeling feasible in many marketing applications.

In addition to the applications already mentioned, there are several major areas in marketing where compound univariate stochastic models have previously been developed, but where copula models can give deeper insight. Examples include purchase incidence and purchase timing, where incidence has a NBD (Morrison and Schmittlein 1988) while inter-purchase timing can be handled with a hazard (Jain and Vilcassim 1991) or exponential model (Schmittlein, Colombo and Morrison 1987). Furthermore, Leeflang et al. (2000, p. 247) list 17 models that integrate purchase incidence, timing and brand choice. All of these models use different distributions for these three components, which somehow need to be combined. Turning attention to the CRM literature, a key concern is whether or not a person in a firm's database is still a "live" customer. Schmittlein et al. (1987) address this issue using a NBD model for purchase occasions compounded with a Pareto model for customer life length. Similarly, for a subscription service, Danaher (2002) develops separate models for cell phone usage and length of time the service is retained. Both these situations can be better accommodated with bivariate models. There is also a large literature on brand choice across two purchase occasions (Lilien, Kotler and Moorthy 1992, pp. 40-53). Here, the marginal distributions are usually the same (perhaps being a Dirichlet-multinomial), but examination is restricted to just two purchases. A multivariate model extending across many

purchases would give improved insights into brand loyalty, for example. Finally, we note that a number of existing multivariate models used previously for marketing applications can be expressed as copula models. For example, the multivariate probit employed by Edwards and Allenby (2003) and the multivariate ordered probit ("cut-point") model of Rossi, Gilula and Allenby (2001) are, in fact, special cases of a "Gaussian copula" model with discrete marginal distributions. Hence, copula models are very general in that they provide both a framework encompassing a number of existing multivariate models, and a practical and flexible mechanism for generating many more.

The paper proceeds as follows. We begin by explaining the basic idea behind copula modeling, including elliptical copula, in the bivariate case. Two bivariate examples, one continuous and one discrete, are used as illustrations. The third section details how to estimate copula models, with particular attention to the discrete case, where we introduce a new efficient Bayesian estimation algorithm. In the fourth section we discuss two high-dimension discrete examples and show how copula models perform better than previously used models. Last, we summarize the benefits of copulas in marketing and show how they might be used in other marketing applications.

2 What is a Copula?

2.1 The Basic Idea

For traditional multivariate distributions, once the parametric distribution function has been selected, the marginals are derived by integration. That is, there is no flexibility for the marginals and they are determined precisely by their parent multivariate distribution (Johnson, Kotz and Balakrishnan 1997; Kotz, Johnson and Balakrishnan, 2000). This usually restricts multivariate distributions to have marginals from the same family; for example, the margins of a multivariate normal are also normal. By contrast, with copula modeling the

starting point is the marginals, which need not be from the same family, and are "glued together" using a copula function.

Consider initially the bivariate case, with two random variables, X_1 and X_2 with distribution functions, respectively, $F_1(X_1)$ and $F_2(X_2)$. It is desired to obtain a bivariate distribution function, $F(X_1, X_2)$, with these two margins. Sklar (1959) proved that there always exists a function C, such that

$$F(X_1 = x_1, X_2 = x_2) = C(F_1(x_1), F_2(x_2)) \,, \tag{2.1}$$

where $C(u_1, u_2)$ is itself a distribution function for a bivariate pair of uniform random variables. Sklar labeled C the "copula function", and showed that it meets three conditions.[1] What is immediately apparent from equation (2.1) is the bivariate distribution function is constructed from the marginals. The role of the copula function is simply to determine the dependence between X_1 and X_2.

If the margins are continuous[2], differentiating equation (2.1) gives the bivariate density for the data

$$f(X_1 = x_1, X_2 = x_2) = c(F_1(x_1), F_2(x_2)) f_1(x_1) f_2(x_2) \,, \tag{2.2}$$

where $c(F_1(x_1), F_2(x_2))$ is called the "copula density", derived by differentiation as $\partial^2 C/(\partial u_1 \partial u_2)$ at $u_1 = F_1(x_1)$ and $u_2 = F_2(x_2)$. Equation (2.2) shows how the copula density controls the level of dependence between X_1 and X_2. For example, if $C(u_1, u_2) = u_1 u_2$, then the copula density is simply $c(u_1, u_2) = 1$, in which case the bivariate density f is just the product of the univariate marginals, so that X_1 and X_2 are independent. Hence, the copula function $C(u_1, u_2) = u_1 u_2$ is known as the "independence copula".

A more general copula function, known as the Farlie-Gumbel-Morgenstern (FGM) copula,

[1] The conditions are given in full in Nelsen (1999; p.45) for the m-dimensional case and are: (i) For every $u \in [0,1]^m$, $C(u) = 0$ if at least one element of u is 0; (ii) For every $u \in [0,1]^m$, $C(u) = u_j$ if all co-ordinates of u are 1, except u_j; and (iii) C is an m-increasing function. In equation (2.1) $m = 2$.

[2] We discuss the case when one or more margins are discrete in Sections 2.3 and 3.1.

(Trivedi and Zimmer 2005, p. 15) is $C(u_1, u_2) = u_1 u_2[1+\tau(1-u_1)(1-u_2)]$. Here, the copula density is $c(F_1(x_1), F_2(x_2)) = 1+\tau(1-2F_1(x_1))(1-2F_2(x_2))$, and so the bivariate density is

$$f(X_1 = x_1, X_2 = x_2) = f_1(x_1) f_2(x_2)[1 + \tau(1 - 2F_1(x_1))(1 - 2F_2(x_2))]\,. \tag{2.3}$$

The parameter τ controls the level of dependence between X_1 and X_2, and the original univariate marginals can still be any distribution. Gumbel (1960) showed that for equation (2.3) to be a density $-1 \leq \tau \leq 1$, which limits the correlation between X_1 and X_2 to the range $(-1/4, 1/4)$.

Equation (2.3) has also been derived via a different class of multivariate distributions, known as Sarmanov distributions (Lee 1996; Sarmanov 1966). It is not difficult to prove that the Sarmanov class of distributions can also be represented as copulas. Sarmanov distributions have appeared before in the marketing literature (see Park and Fader 2004 and Danaher 2007), but they do not extend easily to three or more dimensions. Moreover, as per the FGM copula, the Sarmanov is limited in its ability to model even moderate-sized levels of correlation. Later we will contrast much more flexible copula models than the Sarmanov for two applications and demonstrate the superiority of these copula models.

There are many possible copula functions, the most popular of which are given by Trivedi and Zimmer (2005, p. 16) and Frees and Valdez (1998, p. 25). A major point of difference among them is the range in their correlation coefficients. One type of copula, called the "Gaussian copula", has nearly the full $(-1, 1)$ range in pairwise correlation and is therefore a general and robust copula for most applications (Song 2000). Furthermore, the Gaussian copula has the desirable property that as the number of dimensions, m, increases, the number of parameters in the multivariate density increases only of the order m^2. In comparison, for the Sarmanov, the number of parameters increases by order of 2^m. We discuss this issue in further detail in Sections 2.5 and 2.6.

It is important to understand that the functional form of the copula function does not

determine the distribution of the marginals. For example, a copula function can be based on the Gaussian distribution while one margin has a NBD and the other a gamma distribution. The copula function merely determines the *dependence* between the two random variables, but has no influence on the marginals themselves.

Equation (2.1) can easily be generalized to m dimensions. If the elements of the random vector $X = (X_1, \ldots, X_m)'$ have marginal distributions $F_1(X_1), \ldots, F_m(X_m)$, then the joint distribution function is given by

$$F(X_1 = x_1, \ldots, X_m = x_m) = C(F_1(x_1), \ldots, F_m(x_m)) \,. \tag{2.4}$$

The copula function C now maps $[0,1]^m$ onto $[0,1]$ and still satisfies the conditions noted previously.

The manner in which copula functions are used to model dependency varies depending on whether each X_j is continuous or discrete-valued. Below we outline each case separately and demonstrate them with a bivariate motivating example. For the continuous case we consider duration of visit and purchase amount for buyers at an online retailer, and additionally illustrate a bivariate copula with very different marginal distributions. In the next example we consider category purchases of bacon and eggs in grocery stores, exhibiting how copulas accommodate discrete variables.

2.2 Motivating Example 1: Continuous Margins

One of the more frustrating realities for firms engaged in online commerce is the small conversion rate from visits to sales (Moe and Fader 2004). For example, using data detailed later for amazon.com, only 4.5% of all visits to its website result in eventual sales. Doubtless, conversion rates are even lower for less experienced or well-known online merchants. To combat these low conversion rates, online sellers attempt to make their website more "sticky"

and useful to visitors by offering product information, interactive features, consumer reviews and comparative prices. The rationale is that the longer a visitor is browsing a website, the more likely they are to find what they want or be convinced/enticed into purchasing an item (Bucklin and Sismeiro 2003; Danaher, Mullarkey and Essegaier 2006). Hence, a naïve, but reasonable, starting point is to calculate the dependence between website visit duration and amount purchased.

One of the best-known online retailers is amazon.com, so in this example we use online visit and transaction data made available by comScore for research purposes and sourced from Wharton Research Data Services (www.wrds.upenn.edu). For a panel of 100,000 homes and transactions in the month of September 2002 we have the length of each visit to amazon.com and the transaction amount if a purchase is made. Over 95% of visits result in no purchases, so we restrict ourselves to just those 1442 visits for which a transaction occured. The transaction amounts range from \$1 up to \$2499, while site visits average about 18 minutes when a purchase is made. For these data the empirical Pearson correlation coefficient between visit duration and purchase amount is only 0.08, which is low. A natural conclusion might be that visit duration and purchase amount are not strongly related, so efforts to increase website stickiness may not be worthwhile.

As we shall see, this proves not to be the case. The flaw in the previous approach is the use of the empirical Pearson correlation coefficient as a measure of dependence, as discussed later. It implicitly assumes that the margins are normally distributed, whereas frequently they are not. For instance, Figure 1(a) is a bivariate plot of visit duration against total spend for people making purchases at amazon.com. Clearly, both marginal distributions are highly right-skewed, with a large concentration of data in the lower left-hand position. A variety of parametric models can be fit to each margin. For the duration of website visits (X_1) an appropriate distribution is the log-normal distribution (Danaher et al. 2006). It is a two-parameter distribution, with parameter vector $\theta_1 = (\mu_1, \sigma_1)$, where μ_1 determines the location and σ_1 the scale. A particularly good fit for the sales data (X_2) is obtained using a

generalized extreme value (GEV) distribution.[3] The GEV is a three parameter distribution, with parameter vector $\theta_2 = (k_2, \sigma_2, \mu_2)$, where k_2 determines shape, μ_2 location and σ_2 scale. When fitted to the data using maximum likelihood estimation (MLE) on each univariate margin, the parameter estimates are $\hat{\theta}_1 = (3.3242, 0.9501)$ and $\hat{\theta}_2 = (0.3641, 15.0911, 19.3828)$.

When the margins are continuous, a useful way to think of the copula method is that via the probability integral transformation on each margin, the original random variables X_j are each transformed into uniform random variables $U_j = F_j(X_j)$. That is, no matter what the distribution of the original random variable X_j, the transformed variable U_j is always uniformly distributed. Nevertheless, if the original univariate distributions are dependent, this dependency will carry through to the transformed uniform distributions. The advantage is this dependency is generally easier to capture for the transformed data.

Using the two fitted distributions here, this so-called "copula data" for each margin $j = 1, 2$ can be computed as $u_{ij} = F_j(x_{ij}|\theta_j = \hat{\theta}_j)$, for observations $i = 1, 2, \ldots, n$. Figure 1(b) plots the copula data for visit duration and transaction amount. It is evident that in each margin the univariate distribution is close to uniform on $[0, 1]$. It now remains to capture any dependence between the bivariate copula data. This is achieved by applying a copula function C to these copula data. This will be demonstrated in Section 2.5, where we detail two copula functions that can easily be generalized to higher dimensions. For the moment, we leave the continuous example, returning to it in Section 2.5.1, and now introduce a discrete example.

[3]The GEV distribution has probability density function

$$f(x|\theta) = \left(\frac{1}{\sigma}\right) \exp\left\{-\left(1 + k\frac{(x-\mu)}{\sigma}\right)^{-1/k}\right\} \left(1 + k\frac{(x-\mu)}{\sigma}\right)^{-((k+1)/k)}, \text{ for } x > 0,$$

such that $1 + k(x - \mu)/\sigma > 0$; see Johnson, Kotz and Balakrishnan (1995; pp.75-85).

2.3 Motivating Example 2: Discrete Margins

While many studies in marketing have examined brand choice among alternatives within a product category, there is also interest in looking at purchases across categories. Examples include purchases of diapers and baby food, ground beef and hamburger buns and bacon and eggs. For bacon and eggs, since they are often eaten together, does this translate to them also being purchased together? Danaher and Hardie (2005) looked at this question and we use their reported data to re-examine this question using a copula model.

For discrete data, the steps for copula modeling are similar to that of continuous data, but with one key difference. The initial step is still to use the probability integral transformation of each margin. However, for a discrete distribution, F_j is a step function, and we instead define U_j to be related to the variable X_j through the inequality

$$F_j(X_j - 1) < U_j < F_j(X_j)\,, \tag{2.5}$$

rather than via a direct equality as in the continuous case where F_j is monotonic. Nevertheless, this still ensures U_j is uniformly distributed on $[0, 1]$. Because of this interval based formulation, once distributions are fitted to the margins the copula data can no longer be computed exactly, but instead are considered as latent variables u_{ij}, bounded so that $F_j(x_{ij} - 1) < u_{ij} < F_j(x_{ij})$.

Danaher and Hardie's (2005) data (in their Table 1) are sourced from Information Resources, Inc., a consumer panel based in a large U.S. city; see Bell and Latin (1998) for details. The reported data are the number of times out of four shopping trips when bacon is purchased, and similarly for eggs. Table 1 gives the observed frequencies for the $n = 548$ observations. Danaher and Hardie (2005) find that an appropriate model for each margin is the Beta-Binomial distribution (BBD). The BBD is derived from a binomial distribution for the number of purchases, where the probability of a purchase varies according to a beta distribution, with parameter vector $\theta = (\alpha, \beta)$, to account for consumer heterogeneity. For

the bacon and eggs data the MLEs of the parameters are $\hat{\theta}_1 = (0.8592, 3.9593)$ for bacon, and $\hat{\theta}_2 = (0.3571, 4.4551)$ for eggs. With these fitted BBDs, bounds for the latent copula data u_{ij} are computed using the two distribution functions. Table 2 reports these bounds for the 0 through 4 possible observed purchases in a four-week period for both X_1 and X_2. To complete the picture, note that the highest upper bound $F_j(4|\theta_j = \hat{\theta}_j) = 1$ and lowest lower bound $F_j(-1|\theta_j = \hat{\theta}_j) = 0$.

We return to this example later, after discussing measures of dependence.

2.4 Measuring Dependence

Since the purpose of the copula function is to capture dependence among the univariate marginals, copulas inevitably raise the issue of how to measure dependence (Joe 1997). The most common measure of dependence between pairs of variables is Pearson's correlation, defined for variables X_{j_1} and X_{j_2} as

$$\rho^p_{j_1 j_2} = \mathrm{cov}(X_{j_1}, X_{j_2}) / \sqrt{\mathrm{var}(X_{j_1})\,\mathrm{var}(X_{j_2})}\,. \tag{2.6}$$

In fact, $\rho^p_{j_1 j_2}$ ought to be viewed as a population parameter arising naturally when the variables are normally distributed. The usual estimate of $\rho^p_{j_1 j_2}$ is the empirical correlation coefficient. Use of this correlation coefficient is so widespread that the requirement of normally distributed marginals is commonly overlooked. That is, when the underlying marginal distributions of the data are non-normal, the empirical correlation coefficient is likely to be a poor measure of dependence.

Another measure of dependence for multivariate data is the Spearman rank order correlation coefficient. This measure calculates the Pearson empirical correlation coefficient not for the raw data, but between the ranks of the data. It is rarely appreciated that the rank order correlation coefficient is a sample estimate of Spearman's population correlation measure

(Joe 1997). For continuous margins, this alternative population measure of dependence is defined as

$$\rho^s_{j_1 j_2} = \text{corr}(F_{j_1}(X_{j_1}), F_{j_2}(X_{j_2})). \tag{2.7}$$

This measure is intimately linked with the copula approach to modeling dependence. This is because it is simply the correlation between the probability integral transformed random variables $U_j = F_j(X_j)$, which themselves have the copula function C as their distribution function, as exhibited in equation (2.1). Since U_j is uniformly distributed, it has a known variance of $1/12$ and expectation of $1/2$. Therefore, equation (2.7) can be simplified as

$$\begin{aligned} \rho^s_{j_1 j_2} &= \text{cov}(U_{j_1}, U_{j_2})/(\text{var}(U_{j_1})\,\text{var}(U_{j_2}))^{1/2} \\ &= 12E\left[\left(U_{j_1} - \frac{1}{2}\right)\left(U_{j_2} - \frac{1}{2}\right)\right] \\ &= 12E\left[U_{j_1} U_{j_2}\right] - 3. \end{aligned} \tag{2.8}$$

The covariance of U_{j_1} and U_{j_2} does not depend on the marginal distributions of the original random variables X_{j_1} and X_{j_2}. It is for this reason that Spearman's correlation is not sensitive to the marginal distributions of the raw data, whereas Pearson's correlation assumes normality. Later, we show how a naïve use of Pearson's sample correlation coefficient can mask the true strength of dependence between two random variables.[4]

2.5 Elliptical Copula Functions

After selecting and fitting the marginal distributions, the next stage in the modeling process is to choose a copula function. There are many possibilities, a number of which are given by Joe (1997), Nelsen (2006) and Trivedi and Zimmer (2005). However, the so-called "elliptical copula" have proven the most popular in applied modeling due to the ease with which they can be estimated in dimensions $m \geq 2$; something we discuss in more detail subsequently.

[4]Note that this definition for $\rho^s_{j_1 j_2}$ can also be employed when one or more of the margins are discrete by uisng the definition for U_j in equation (2.5).

Here, the copula function is based on an elliptical distribution, such as the multivariate t or Gaussian, but should not be confused with using an elliptical distribution for the data itself. The key idea is to transform the uniformly distributed random vector $U = (U_1, \ldots, U_m)'$ to another random vector $X^\star = (X_1^\star, \ldots, X_m^\star)' \in \mathcal{R}^m$ and then model $X^\star$ using an elliptical distribution to capture the dependence among the elements of $X^\star$, and therefore also the original data X.

The simplest elliptical copula function is the Gaussian, which corresponds to the following transformation. If $X_j^\star = \Phi^{-1}(U_j)$, with Φ the univariate standard normal distribution function, then the vector $X^\star$ is modeled as $N(0, \Gamma)$, where Γ is a correlation matrix. This transformation corresponds to using the following Gaussian copula function for C in equation (2.4):

$$\begin{aligned} C_G(u) &= \Phi_m(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_m)|\Gamma) \\ &= \Phi_m(x_1^\star, x_2^\star, \ldots, x_m^\star|\Gamma)\,, \end{aligned} \tag{2.9}$$

where $\Phi_m(\cdot|\Gamma)$ is the distribution function of a multivariate $N(0, \Gamma)$ distribution and we define $x_j^\star = \Phi^{-1}(u_j)$. The correlation matrix Γ captures dependence among the elements of $X^\star$, and therefore also among the elements of the vector X.

Another popular elliptical copula in the finance literature is based on the multivariate t distribution (Daul et al., 2003), which corresponds to the following transformation. Let $X_j^\star = T_\nu^{-1}(U_j)$, with T_ν the distribution function of a univariate student t distribution with ν degrees of freedom, then $X^\star$ is modeled as multivariate t with ν degrees of freedom, mean 0 and scale matrix Γ_T, where as in the Gaussian case, Γ_T is a correlation matrix. This corresponds to assuming the following t-copula function for C in equation (2.4):

$$\begin{aligned} C_T(u) &= T_{m,\nu}(T_\nu^{-1}(u_1), T_\nu^{-1}(u_2), \ldots, T_\nu^{-1}(u_m)|\Gamma_T) \\ &= T_{m,\nu}(x_1^\star, x_2^\star, \ldots, x_m^\star|\Gamma_T)\,. \end{aligned} \tag{2.10}$$

Here, $T_{m,\nu}(\cdot|\Gamma_T)$ is the distribution function of a multivariate t distribution with location 0, degrees of freedom ν and $m \times m$ scale matrix Γ_T, and $x_j^\star = T_\nu^{-1}(u_j)$. As with the Gaussian copula function, the correlation matrix Γ_T captures the level of dependence in $X^\star$, and therefore also X.

When applied to data, both these elliptical copula correspond to a transformation of the copula data for each margin u_{ij} to new values $x_{ij}^\star$. It is these transformed copula data that are used to fit either a zero-mean multivariate normal or t distribution.

2.5.1 Motivating Example 1 Continued

We now return to the previous website duration and purchase amount example to illustrate the inverse-normal transformation and Gaussian copula function when the margins are continuous. Figure 1(c) plots the transformed copula data $x_{ij}^\star = \Phi^{-1}(u_{ij})$, $i = 1, \ldots, n$, for the website visit duration ($j = 1$) and spend ($j = 2$) data. A mild positive dependence is apparent visually. This dependence is measured by the estimated off-diagonal element γ_{12} of Γ when fitting a bivariate $N(0, \Gamma)$ distribution to the twice-transformed data in Figure 1(c). Using maximum likelihood estimation for this continuous example, as outlined later in Section 3.1, the estimated off-diagonal element is $\hat{\gamma}_{12} = 0.233$. The estimated Gaussian copula, along with fitted GEV and log-normal marginal distributions, fully define the joint distribution of the random vector X; a distribution that is parametric.

Figure 1(d) plots the transformed copula data using the inverse t distribution function, $x_{ij}^\star = T_5^{-1}(u_{ij})$ with $\nu = 5$ degrees of freedom.[5] A t_5-copula corresponds to fitting a multivariate $t_{m,5}(0, \Gamma_T)$ distribution to this transformed data, with the level of dependence being parameterized by the off-diagonal element of Γ_T. Again, using maximum likelihood, this is estimated from the data as $\hat{\gamma}_{12} = 0.1880$. This estimated t_5-copula, along with the fitted margins, also fully define a parametric joint distribution for X.

[5] We pick $\nu = 5$ degrees of freedom here to ensure that the t-copula has much heavier tails than a Gaussian. The degrees of freedom can also be estimated from the data, although we do not do so here.

Using the two fitted copula models the question of how much dependence exists between duration and spend can be answered. Table 3 reports several measures of dependence. The first is the empirical Pearson correlation coefficient, which is rather low at 0.08. Following this is the empirical Spearman rank order correlation, which is about triple Pearson's correlation, at 0.26. The Spearman correlation is matched exactly by the two parametric-based Spearman correlations, where either Gaussian- or t-copulas are used in equation (2.7). This indicates there is a moderate correlation, whereas the very low empirical correlation suggests there is almost no association between visit duration and amount spent. There are two distinct reasons for the difference. First, Pearson's correlation measure is really only appropriate when data are approximately normally distributed. Figure 1(a) clearly shows that website visit duration and transaction amount are not normally distributed. Second, the regular correlation coefficient is a nonparametric estimate, and is likely to be inferior to an estimate obtained from a well-fitted parametric distribution, such as the Gaussian copula.

To further demonstrate the usefulness of having a parametric bivariate distribution, we calculate the expectation of total spend, conditional on visit duration, for the Gaussian copula model. That is, $E[X_2|X_1]$, which is directly obtainable from the bivariate distribution.[6] Figure 1(e) plots this conditional expectation for different duration levels ranging from the lowest to highest quintiles in our data. It reveals that the expected spend increases substantially (with diminishing returns) as visit duration increases - in fact, more than doubling from \$28.76 to \$58.21 over this range of duration. Hence, the previous naïve analysis using just the empirical correlation misses the subtle, but substantial, relationship between website visit duration and purchase amount. This demonstrates that copula modeling can reveal previously undetected dependences and can give managers improved insights into relationships among variables in their databases.

[6] We compute $E[X_2|X_1 = x_1] = \int x_2 f(x_1, x_2) \mathrm{d}x_2$ using numerical integration in Matlab.

2.5.2 Motivating Example 2 Continued

We now return to the bacon and eggs example, where the marginals are discrete BBDs. When the margins are discrete-valued the bounds on U_j are transformed to obtain new bounds for the random variable $X_j^\star$. For example, for a Gaussian copula the bounds on $X_j^\star = \Phi^{-1}(U_j)$ are

$$\Phi^{-1}(F_j(X_j - 1)) < X_j^\star < \Phi^{-1}(F_j(X_j))\,. \tag{2.11}$$

Correspondingly, the latent copula data $u_i = (u_{i1}, \ldots, u_{im})$ are transformed to a second set of latent variables $x_i^\star = (x_{i1}^\star, \ldots, x_{im}^\star)'$ which are distributed $N(0, \Gamma)$ but with bounds $\Phi^{-1}(F_j(x_{ij} - 1)) < x_{ij}^\star < \Phi^{-1}(F_j(x_{ij}))$. To illustrate with the bacon and eggs data, Table 2 also contains the bounds on these second latent variables, where there is one bound corresponding to each discrete value in the data. Using the Bayesian method of estimation outlined later in Section 3.2, the off-diagonal element of the matrix Γ for the Gaussian copula function is $\hat{\gamma}_{12} = 0.2895$.

From the raw data, the empirical Pearson correlation coefficient is 0.233, and the Spearman rank correlation coefficient is 0.217. This indicates a moderate correlation between bacon and egg purchases, so they are often purchased together as well as eaten together. However, each of these metrics are unreliable - the empirical Pearson correlation coefficient because the data are very far from normally distributed, and the Spearman rank correlation coefficient because the data are discrete and there are many tied ranks. However, using the Gaussian copula we can compute a reliable estimate of the Spearman correlation $\rho_{j_1 j_2}^s = 0.286$ defined in equation (2.8) using the Bayesian method outlined below in Section 3.3. This slightly higher level of dependence confirms the higher level of dependence also uncovered using the Sarmanov copula by Danaher and Hardie (2005).

2.6 Advantages of Elliptical Copulas

As demonstrated in the bivariate examples above, each margin can be modeled separately and multivariate dependence added at a later step through the choice of copula function. Elliptical copulas are often preferred because they have three major advantages over alternative choices. First, dependence in the data is parameterized by a correlation matrix Γ which has only $m(m-1)/2$ unique elements. Most other copula functions, including the Sarmanov, prove difficult to extend to higher dimensions - something we explore later in Section 4.

Second, random iterates can be generated from the joint density F both quickly and simply. Below we outline an algorithm to generate an iterate X from a Gaussian copula with arbitrary marginal distributions $F_1, \ldots, F_m$:

Algorithm 1: Random Iterate Generation

Step 1: Generate $Z \sim N(0, \Gamma)$

Step 2: Set $U = (\Phi(Z_1), \ldots, \Phi(Z_m))'$

Step 3: Set $X = (F_1^{-1}(U_1), \ldots, F_m^{-1}(U_m))'$

This algorithm can be easily adjusted for other elliptical copula by replacing the Gaussian distribution in Steps 1 and 2.[7] The ability to simulate from an elliptical copula is particularly useful for evaluating the distribution of any summary or metric based on X using Monte Carlo simulation. It is in this manner that we compute estimates of $\rho^s_{j_1 j_2}$ and $\rho^p_{j_1 j_2}$ in Table 3. In Section 4 we also use this Monte Carlo algorithm to evaluate the distribution of the summation of the elements of X.

Third, an elliptical copula function allows for the interpretation of the copula model for the distribution F as a transformation from X to the elliptically distributed $X^\star$. When one or more of the margins are discrete-valued, $X^\star$ is a bounded latent variable. For example, the multivariate probit model used by Edwards and Allenby (2003) is one example of a

[7] For example, for the t_5 copula, Step 1 would have $Z \sim t_{m,5}(0, \Gamma)$ and Step 2 has $U = (T_5(Z_1), \ldots, T_5(Z_m))'$.

Gaussian copula. Here, the latent variables are jointly normally distributed, and the margins are binary-valued with differing distributions F_j determined by m univariate probit models. In Section 3.2 we exploit the latent variable representation to propose a Bayesian approach to estimation where the realizations $\{x_1^\star, \ldots, x_m^\star\}$ of the latent variable are explicitly generated in a Markov chain Monte Carlo (MCMC) algorithm. This extends the widely used MCMC method of estimation for choice models, where realizations of the latent variables are explicitly generated (Albert and Chib 1993; Edwards and Allenby 2003).

To demonstrate the flexibility of the approach we use Algorithm 1 to generate 50,000 iterates from the Gaussian copula fitted to the bacon and eggs data. From these, we can compute the probability mass function, which is reported in Table 1. Also reported is the mass function from a fit using the same BBD marginals and the Sarmanov copula as given by Danaher and Hardie (2005). There is closer agreement between the fitted Gaussian copula and the empirical distribution in Table 1. Indeed, the sum of the absolute differences between the empirical and fitted probabilities is 0.0751 for the Sarmanov and 0.0625 for the Gaussian.

3 Estimation of Copula Models

In a copula model there are two sets of parameters that require estimation. The first set is the parameters of each of the selected marginal distributions, $\Theta = \{\theta_1, \ldots, \theta_m\}$, and the second is the dependence parameters of the selected copula function. When an elliptical copula is employed, the latter is the $m(m-1)/2$ non-fixed parameters in the correlation matrix Γ. The most common approach is to use a two-stage estimation procedure, where the parameters for the margins are estimated separately, and then Γ is estimated conditional upon these. The estimation of the marginal parameters in the first stage is usually straightforward and can be performed in a wide variety of ways, including maximum likelihood, Bayesian, or a method of moments based approach. However, the estimation of Γ is more difficult, and is different depending on whether the variables $X_1, \ldots, X_m$ are continuous or discrete. We

deal first with the continuous case, where parameters can be estimated reliably in most cases using maximum likelihood.

3.1 Maximum Likelihood Estimation

If $x = \{x_1, \ldots, x_n\}$ are n observations on m continuous margins, then the likelihood is $L(\Theta, \Gamma; x) = \prod_{i=1}^{n} f(x_i|\Theta, \Gamma)$, where f is the density function derived from the copula model in equation (2.4). As for the bivariate case in equation (2.2), the contribution of the ith observation to L is

$$f(x_i|\Theta, \Gamma) = c(F_1(x_{i1}), \ldots, F_m(x_{im})) \prod_{j=1}^{m} f_j(x_{ij}) \,, \tag{3.1}$$

where the vector $x_i = (x_{i1}, \ldots, x_{im})'$ and c is the copula density. For example, following Song (2000), for the Gaussian copula the copula density is

$$\begin{aligned} c(u_{i1}, \ldots, u_{im}) &= \frac{\partial C(u_{i1}, \ldots, u_{im})}{\partial u_{i1} \cdots \partial u_{im}} \\ &= |\Gamma|^{-1/2} \exp\left\{ -\frac{1}{2} x_i^{\star\prime} (\Gamma^{-1} - I) x_i^{\star} \right\} . \end{aligned}$$

Recall from Section 2.5 that $x_i^{\star}$ is the observation x_i twice-transformed, with the jth element being $x_{ij}^{\star} = \Phi^{-1}(F_j(x_{ij}))$. Because this uses the probability integral transformations based on the marginal distributions, $x_i^{\star}$, and therefore f, are functions of the marginal parameters Θ. For the two-stage estimator, maximization of the log-likelihood $l(\Theta, \Gamma) = \log(L(\Theta, \Gamma|x))$ can be undertaken numerically with respect to Γ. For full maximum likelihood estimation the resulting point estimates can be used as the initial conditions for maximization of l with respect to all the parameters $\{\Theta, \Gamma\}$, although the resulting estimates usually differ only slightly.

The likelihood can also be derived for the discrete case, but this is more involved and the resulting expression difficult to actually compute. For a discrete distribution,

to derive the density f of each observation, the so called Radon-Nikodym derivative of $F(x) = C(F(x_1), \ldots, F_m(x_m))$ has to be taken with respect to $x = (x_1, \ldots, x_m)$, which is a discrete-valued measure. Details of this can be found in Song (2000), and it results in the expression

$$f(x) = P(X_1 = x_1, \ldots, X_m = x_m) = \sum_{k_1=1}^{2} \cdots \sum_{k_m=1}^{2} (-1)^{(k_1+\ldots+k_m)} C(\tilde{u}_{1k_1}, \ldots, \tilde{u}_{mk_m}), \quad (3.2)$$

where $\tilde{u}_{j1} = F_j(x_j)$, and $\tilde{u}_{j2} = F_j(x_j-)$ is the left-hand limit of F_j at x_j (Trivedi and Zimmer 2005, p.56). Unfortunately, there are severe problems in computing and optimizing the resulting log-likelihood. First, there are 2^m terms in the sum in equation (3.2), so that evaluating them all can be impractical even for moderate values of m. Second, when an elliptical copula is employed, m-dimensional multivariate distribution functions have to be evaluated to compute each term in the sum. While there have been a number of advances in techniques to undertake this (see, for example, Genz 1992 and Genz and Bretz 2002) this still remains a difficult problem involving significant numerical error and computation burden. What's more, this has to be repeated $n2^m$ times to evaluate the likelihood just once.

Overall, the computational demands in the discrete case are prohibitive even for a moderate number of dimensions. To compound the problem further, the log-likelihood can prove difficult to optimize for many choices of copula function, even when the dimension is as low as $m = 3$ (Trivedi and Zimmer 2005). These problems with implementing maximum likelihood estimation have hindered the adoption of copula modeling for problems with one or more discrete-valued margins; precisely the types of problems that often arise in marketing. However, Pitt et al. (2006) recently outline a straightforward, flexible and general Bayesian approach for the estimation of Gaussian copula models when the margins are discrete, continuous or mixed. This approach shows great promise for the analysis of marketing data, so we now outline and extend their method.

3.2 Bayesian Simulation Solution

Over the past fifteen years Bayesian estimation, where parameter estimates are obtained from their posterior distribution, have become increasingly popular. The approach has proven particularly useful for more complex and high-dimensional models where there can be many parameters. Here, so called Markov chain Monte Carlo (MCMC) simulation algorithms are used to generate a Monte Carlo sample from the posterior distribution. It is from this Monte Carlo sample that posterior estimates and other inference is computed; see Robert and Casella (2004) for an accessible introduction to MCMC estimation. A major advantage of MCMC simulation is that each of the parameters can be generated conditional on the other parameters in the model, making each step of the sampling scheme relatively easy to implement for complex models. Such approaches have also had an impact in the marketing literature; see, for example, Bradlow and Rao (2000), Smith, Mathur and Kohn (2000), Rossi and Allenby (2003), and Rossi, Allenby and McCulloch (2005).

Pitt et al. (2006) propose using an MCMC solution for the Gaussian copula model. They point out that if the jth margin is discrete-valued, the problem of estimation can be greatly simplified by treating $x^\star_{ij}$, for $i = 1, \ldots, n$, as latent variables. For the Gaussian copula these latent variables can be generated explicitly in the MCMC scheme from constrained Gaussian distributions. The correlation matrix Γ of the Gaussian copula can be generated conditional on $x^\star = \{x^\star_{ij}\,; i = 1, \ldots, n;\, j = 1, \ldots, m\}$ at each sweep of the simulation algorithm. The following MCMC algorithm is based on that found in Pitt et al. (2006):

Algorithm 2: Bayesian MCMC Simulation Algorithm

Step 1: For $j = 1, ..., m$:

Step 1(a): If margin j is continuous, set $x^\star_{ij} = \Phi^{-1}(F_j(x_{ij}))$, for $i = 1, \ldots, n$.

Step 1(b): If margin j is discrete, generate from the conditional distribution

$$x^\star_{ij} | \{x^\star \backslash x^\star_{ij}\}, \Gamma, x, \text{ for } i = 1, \ldots, n.$$

Step 2: Generate from the conditional distribution $\Gamma | x^\star, x$.

Here, $\{x^\star \backslash x^\star_{ij}\}$ denotes all the values of $x^\star$, excluding the element $x^\star_{ij}$, while x is the vector of observed data. One repetition of the algorithm is called a "sweep" in the MCMC literature, and the approach requires many repeated sweeps. This algorithm allows estimation of the Gaussian copula, but is extendable to other elliptical copula.

Step 1(b) involves generating the latent variables one element, $x^\star_{ij}$, at a time from its Bayesian conditional posterior distribution. The appendix shows how to derive this distribution, which is a univariate constrained distribution, and facilitates fast generation. There are nm such latent variables, so that the computational demand of this step increases only linearly with both dimension and sample size, making it practical for even reasonably large problems. In Section 4 we demonstrate this by applying the approach to a problem with $m = 45$ dimensions and $n = 10000$ observations.

Step 2 requires generation of the correlation matrix Γ from its posterior distribution, conditional on values for $x^\star$. This is the most difficult part of the sampling scheme because generating a correlation matrix is a challenging statistical problem (Bernard, McCulloch and Meng 2000). Pitt et al. (2006) present a method to generate Γ based on a complex prior, which is difficult to both interpret and implement. Instead, we propose a simpler alternative that is based on the random walk Metropolis-Hastings algorithm (see Robert and Cassella 2004, pp. 287-291, for a discussion of this technique) and the following representation of Γ:

$$\Gamma = \text{diag}(\Sigma)^{-1/2}\, \Sigma\, \text{diag}(\Sigma)^{-1/2}\,. \tag{3.3}$$

Here, Σ is a non-unique positive definite matrix and $\text{diag}(\Sigma)$ is a diagonal matrix comprised of the leading diagonal of Σ. We further decompose $\Sigma^{-1} = R'R$, with R being an upper triangular Cholesky factor. If we set the leading diagonal of R to all ones, this leaves $m(m-1)/2$ non-fixed elements of R, matching the number of non-fixed elements of Γ and identifying the representation. The advantage of this seemingly round-about representation is that upper triangular elements of R are unconstrained and can be generated easily one

element at a time using a random walk Metropolis-Hastings step. Regardless of the values for R we generate, the transformation ensures that Γ remains a postive definite correlation matrix. Similar transformations have been employed in estimating covariance matrices in longitudinal models (Smith and Kohn 2002; Panagiotelis and Smith 2008).

Using this representation, we propose the following algorithm to generate values of Γ at Step 2 of Algorithm 2:

Algorithm 3: Generation of Γ

Step 2: Repeat the following for $i = 1, \ldots, m-1$ and $j > i$:

Step 2(a): Generate r_{ij} (elements of R) using a random walk Metropolis-Hastings step.

Step 2(b): Compute $\Sigma = (R'R)^{-1}$.

Step 2(c): Compute $\Gamma = \text{diag}(\Sigma)^{-1/2}\,\Sigma\,\text{diag}(\Sigma)^{-1/2}$.

Generation of Γ in this fashion proves to be quick, reliable and scales up to higher dimensions well. The appendix provides further details on how to implement both Algorithms 2 and 3. Once run for an initial period, Algorithm 2 provides a Monte Carlo sample of K iterates, which we denote as $\{(\Gamma^{[1]}, x^{\star\,[1]}), \ldots, (\Gamma^{[K]}, x^{\star\,[K]})\}$ with the superscript here labeling the iterate number. The sample can be shown to be distributed $\Gamma, x^\star|x$, which is the Bayesian posterior distribution of Γ augmented with the latent variables $x^\star$, conditional on the observed data x. It is from this output of Algorithm 2 that Bayesian estimates are computed.

3.3 Bayesian Estimates

Bayesian point estimates of parameters and other metrics are usually given by their posterior means. The parameter estimates can be computed directly from the Monte Carlo sample output from Algorithm 2. For example, the posterior mean of the correlation matrix of the

elliptical copula is:

$$E(\Gamma|x) \approx \frac{1}{K}\sum_{k=1}^{K}\Gamma^{[k]}\,. \tag{3.4}$$

Another useful aspect of using MCMC for estimation is that iterates of U and X can be obtained at each sweep by appending Algorithm 1 to the end of Algorithm 2. Again, using the superscript notation to label the iterates, the resulting Monte Carlo sample of K iterates is $\{(U^{[1]}, X^{[1]}), \ldots, (U^{[K]}, X^{[K]})\}$. These iterates are from the fitted distribution with Γ effectively integrated out, thereby removing the uncertainty associated with the estimation of Γ. This is a major improvement over maximum likelihood estimation, where Algorithm 1 can be used to generate realizations from the copula model, but only conditional on the maximum likelihood point estimate of Γ and ignoring any uncertainty associated with the estimate. We shall see in subsequent examples that this can lead to substantial overall improvement in the quality of the resulting estimates.

Using these iterates the posterior distribution of other metrics can be computed. This includes the posterior mean of Spearman's pairwise correlation measure, which can be computed as

$$E(\rho^{s}_{j_1 j_2}|x) = 12E(U_{j_1}U_{j_2}|x) - 3 \approx \frac{12}{K}\sum_{k=1}^{K} U^{[k]}_{j_1}U^{[k]}_{j_2} - 3\,. \tag{3.5}$$

This Bayesian estimate is based on the parametric assumption of an elliptical copula model for the distribution of X, and can differ from the empirical rank correlation coefficient.

Last, we note that other key metrics can also be obtained from the Monte Carlo iterates, depending on the nature of the application. For example, in Section 4 we also use the Monte Carlo iterates $\{X^{[1]}, \ldots, X^{[K]}\}$ to compute the distribution of total advertising exposures in print media and website page views examples.

4 Multivariate Examples

The two examples presented in Section 2 are both bivariate, with Example 1 demonstrating the ability of copulas to link together different marginal distributions. The examples in this section illustrate the power of copula modeling in higher dimensions where the marginal distributions are discrete. This is a situation where the versatility of the Bayesian estimation method also becomes apparent. The first example is for multiple magazines in an advertising campaign, while the second is for page views across many websites. Both examples show that copula models contrast very favorably with previous models in terms of estimation accuracy and holdout validity. Furthermore, the second example illustrates the ability of copulas to deal with a very high number of dimensions.

4.1 Example 3: Magazine Advertising Campaigns

Expenditure on magazine advertising in the U.S. exceeds $12.3 billion annually (Ad Age 2006). There is a long history of models being used to estimate exposure to magazine advertisements (for example, see Chandon 1986; Rust 1986; Danaher 1992). As a medium, magazines pose two modeling challenges. The first is that many readers subscribe to their preferred magazines, giving rise to high intra-magazine correlation. Here, reading an issue of a magazine raises the probability of reading the next and subsequent issues. The second challenge is that people often read several magazines within a genre, such as women's magazines, sports or news. This results in inter-magazine correlation, meaning that exposure to advertisements across multiple magazines cannot be assumed to be independent (Danaher 1992). The first challenge is usually tackled by using a Beta Binomial distribution (BBD) for each magazine's marginal distribution (Chandon 1986; Rust 1986), while the second challenge requires a multivariate model for all the magazines (Danaher 1992), which we now discuss.

Modeling magazine exposure requires the modeling of X_j, which denotes the number

of exposures a person has to magazine j, for $j = 1, 2, \ldots, m$. Usually, advertisers are not so much interested in a full multivariate distribution of $X = (X_1, X_2, \ldots, X_m)$, but rather a function of X. The most common function is total exposure across all magazines in an advertising campaign, denoted here as $S = \sum_{j=1}^{m} X_j$. Even though the managerial requirement is just a sum of the exposures, the full multivariate model is necessary in the first instance due to inter-magazine correlation (Danaher 1992). Models which capture the multivariate distribution first, then use this distribution to estimate the total exposures, always dominate simpler models that directly estimate total exposures (Danaher 1991; 1992). Knowing the distribution of S also enables estimation of key advertising metrics, such as reach, $\Pr(S \geq 1)$, average frequency, $E[S]$/reach, and the frequency distribution, $\Pr(S \geq 1)$, $\Pr(S \geq 2), \Pr(S \geq 3), \ldots$, and so on (Rust 1986; Danaher 1992).

To date, the class of models shown to be best at modeling multivariate magazine exposure are based on "canonical expansions" (Danaher 1991). Danaher and Hardie (2005) show that the canonical expansion model is the same as the Sarmanov model mentioned earlier. The Sarmanov model for multivariate media exposure distributions is

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_m = x_m) = f(x_1, \ldots, x_m) = \prod_{j=1}^{m} f_j(x_j) \times \left[1 + \sum_{j_1 < j_2} \left\{ \omega_{j_1, j_2} \phi_{j_1}(x_{j_1}) \phi_{j_2}(x_{j_2}) + \ldots + \omega_{1,2,\ldots,m} \prod \phi_j(x_j) \right\}\right], \quad (4.1)$$

where $f_j(X_j)$ is the univariate exposure distribution for magazine j (usually a BBD), ω_{j_1,j_2}, $\omega_{j_1,j_2,j_3}, \ldots, \omega_{1,2,\ldots m}$ are bivariate, trivariate, and higher order association parameters, and $\phi_j(x_j)$ are called "mixing functions" with the property that $\sum_{x_j=0}^{\infty} \phi_j(x_j) f_j(X_j) = 0$. As noted earlier, the Sarmanov distribution is nested within the larger class of all copula models, with the Sarmanov copula function in equation (4.1) being the FGM copula.

What is immediately apparent from equation (4.1) is that the number of terms and parameters increases rapidly as the number of magazines, m, increases. In fact, the number

of parameters increases by order 2^m. In response to this cripling computational problem, Danaher (1991) truncated equation (4.1) after just second-order terms, producing an approximate Sarmanov model. While this reduces the computational difficulties, it introduces another problem; namely, that the distribution is no longer well-defined and that the modeled probabilities can be negative. In contrast, an elliptical copula with correlation matrix Γ has only $m(m-1)/2$ parameters, which increases by order m^2 rather than 2^m. As a consequence, there is no need to approximate an elliptical copula model, meaning that the resulting distribution is well-defined and the estimated probabilities are always nonnegative.

The data for this example come from Nielsen Media Research, who regularly conduct fully national surveys of people in many countries and ask them about their magazine reading behavior over the past 6 months. The data employed here come from a market which was surveyed throughout 2007, and the sample size is $12,000$ people. The data used to fit the Sarmanov and Gaussian copula models come from questions which ask respondents how many of the past 4 issues of a magazine they have read. We selected just 10 of the more than 200 magazines included in the survey. The magazines are *New Idea, Listener, Woman's Weekly, Time, TV Guide, Womans Day, North and South, Readers Digest, Cuisine* and *National Geographic*. The women's magazines in this list are particularly highly correlated, with the highest empirical Pearson correlation between any two magazines being 0.59.

As mentioned earlier, a particularly robust model for univariate magazine exposure models is the BBD. We therefore fit separate BBDs to each magazine. As the number of magazines exceeds Danaher's (1991) recommended limit of 6, we approximate the Sarmanov/canonical expansion model to use terms only up to bivariate pairs in equation (4.1). The bivariate association parameters ω_{j_1,j_2} are estimated using pairwise correlations, as shown by Danaher (1991) and Danaher and Hardie (2005). Once the full multivariate distribution is estimated, the probability mass function of the sum f_S can be easily obtained

by summing over these probabilities, so that

$$f_S(s) = \sum_{\{(x_1,\ldots,x_m):x_1+\ldots+x_m=s\}} f(x_1, x_2, \ldots, x_m). \tag{4.2}$$

We also employ a Gaussian copula model with the same maximum likelihood estimates for the BBD margins. The off-diagonal correlations in Γ that define the copula function and the joint distribution of X are estimated by employing Algorithm 2. This produces a Monte Carlo sample from which Γ can be estimated by its Bayesian posterior mean using equation (3.4).

Following the discussion in Section 3.3, Algorithm 1 can be appended to the end of each sweep of Algorithm 2 to provide a Monte Carlo sample of iterates of X which are distributed as the fitted Gaussian copula model. Computing the sum of the elements of each iterate converts this into a Monte Carlo sample of S, from which the relative frequencies provide a Bayesian estimate of the total exposure distribution f_S.

We can compare these estimated parametric distributions with the empirical distributions obtained by summing the total number of exposures across the ten magazines for each person in the data. There is a well-established history of validating magazine exposure models in this way (Leckenby and Kishi 1984; Danaher 1991). Model estimation accuracy is assessed by comparing the estimated probabilities with those from the empirical distribution. Here we use two measures previously employed by Leckenby and Kishi (1984) and Danaher (1991). Denote $f_s = f(S = s)$ and $\hat{f}_s$ as the observed and estimated exposure distribution probabilities, respectively. Then define the relative error in reach (RER) as

$$RER = \frac{|\hat{f}_0 - f_0|}{1 - f_0}$$

and the error in exposure probabilities over reach (EPOR) as

$$EPOR = \frac{\sum_{s=0}^{20} |\hat{f}_s - f_s|}{1 - f_0}.$$

We limit the EPOR calculation to 20 exposures as there is little managerial interest in exposures beyond this range (Rossiter and Danaher 1998). The left hand portion of Table 4 contains the observed distribution and the estimated Sarmanov and Gaussian copula exposure distributions using the full year of data. Also included for comparison is the distribution of total exposure when magazines are treated as independent. The right hand portion of Table 4 facilitates a validation test, whereby the data for the 6000 people interviewed in the first half of the year are used to fit the model parameters, then the fitted model is used to predict the exposure distribution for the second half of the year.

The RER and $EPOR$ statistics show that the copula model is substantially more accurate at estimating the reach and exposure distributon for the full year, and in the validation half-year. They also show that ignoring the inter-magazine dependence results in a very poor estimate of the distribution of total exposure. Moreover, the copula model is substantially quicker, taking 55 mins on a 2.66GHz PC, while the Sarmanov model took 6.5 hours.

There are two reasons for the superior performance of the Gaussian copula model in comparison to the Sarmanov model. The first, is that the Gaussian copula is a well-defined distribution, whereas the Sarmanov is not because of the computational necessity of truncating the sum in equation (4.1). The second is that, as discussed in Section 3.3, the Bayesian method integrates out the inter-magazine dependency structure (parameterised by the Γ matrix), as opposed to conditioning on its point estimate as with the method of estimation for the Sarmanov outlined in Danaher and Hardie (2005). Given that there is some meaningful uncertainty regarding the exact values of Γ, accounting for that uncertainty results in a more accurate estimate of the exposure distribution.

We further demonstrate the usefulness of copulas in this example by showing how the

estimates of the joint distribution can be used to maximize reach. Rust (1986) discusses a number of media optimization methods, but one which is computationally attractive is a heuristic called the "greedy algorithm", which procceeds in a stepwise fashion as follows. First, select the magazine which maximizes reach subject to an overall budget constraint. Next, choose a second magazine which results in the greatest increment in reach conditional on the first magazine already selected, again staying within the budget. This sequence continues until the budget is exhausted. This algorithm dovetails nicely with the MCMC approach, where the Monte Carlo iterates of $X = (X_1, X_2, \ldots, X_m)$ can be used to compute the required conditional distributions. We implemented this greedy algorithm using realistic costs per magazine and a total budget of \$75,000. Using the first 6 months of data three magazines were chosen (*Woman's Weekly, TV Guide, Cuisine*), giving a maximum predicted reach of 62.3%. We tested the accuracy of this heuristic by also conducting a complete enumeration of all the possible magazine combinations, using the empirical reach derived from the observed exposure distribution. This (slow) method gives the exact same magazine combination as the heuristic, with the empirical reach being 62.0%. Lastly, we also undertook a complete enumeration in the validation period, and again the same three magazines comprise the optimal media scedule. This further illustrates the usefulness of copula models with Bayesian estimation, as the MCMC iterates facilitate simple estimation of conditional distributions.

4.2 Example 4: Website Page Views

Our final example returns to the Internet, but in a different context. Not only has the Internet exploded as a sales channel, but in recent years it has become an accepted and essential advertising medium, as evidenced by 30% annual growth since 2003, to the point where online ad spend in the U.S. exceeded \$21 billion in 2007 (IAB 2008). One key difference between the Internet and all other media is the enormous number of potential websites in

which to advertise. This underscores the need for models which are capable of handling many websites. In this example we demonstrate a copula model for up to 45 websites, which substantially exceeds the 15 website maximum modeled previously using the Sarmanov by Danaher (2007).

We follow Danaher (2007) and model page views for multiple websites, since banner ads and other forms of online display ads are delivered via web pages.[8] Let X_j be the number of page views to website j, $j = 1, 2, \ldots, m$. An appropriate and robust model for univariate page view distributions is the NBD (Danaher 2007; Huang and Lin 2006). To date, the best nonproprietary multivariate model for page views is based on a Sarmanov distribution (Danaher 2007). However, as already noted, the full model becomes computationally infeasible for even moderate dimensions, so we follow Danaher (2007) and truncate after trivariate terms. As a result the Sarmanov model runs the risk of producing negative probability estimates. Nonetheless, Danaher (2007) tests the Sarmanov model across a range of popular websites and shows that it outperforms several other possible models in an extensive validation test.

Our copula model for multivariate website page views is similar in principle to the copula model for multivariate BBDs demonstrated above for magazines, except the univariate distributions are now NBDs. To link the univariate NBDs we use the Gaussian copula function as in Example 3. As discussed previously, this has the attractive property of only requiring estimates of $m(m-1)/2$ off-diagonal element for the correlation matrix Γ, resulting in a managable 990 parameters here for all $m = 45$ websites. By contrast, even the truncated Sarmanov model has $15,180$ parameters.

To facilitate a fair comparison, we use comScore data for the United States from the same source as Danaher (2007); namely, the Wharton Research Data Service, where we also sourced the transaction data for our first example. Data for the month of September 2002 are used for estimation and November 2002 for validation, with sub-panels of $10,000$

[8] We do not consider paid search advertising in this example. Display ads are still the dominant form of Internet advertising, being 45% of all Internet ad spend (IAB 2008).

randomly chosen homes for each time period taken from the entire comScore panel of 100,000. There is no overlap in homes across the two sub-panels. The same 45 websites listed in Danaher's (2007) Table 2 were selected, being the most popular sites at that time in the U.S. We fit NBDs to all 45 marginal distributions, and then use MCMC to estimate the Gaussian copula model once for all 45 sites.

We estimate the total exposure distribution in the same manner as in the previous example. That is, we append Algorithm 1 to the end of the Algorithm 2 and generate a large Monte Carlo sample of X, which is a vector with 45 elements. We also generate the corresponding Monte Carlo sample of U, and use these to compute estimates of all 990 possible pairwise Spearman correlations $\rho^s_{j_1 j_2}$ using the Bayesian estimator in equation (3.5). Figure 2 plots these Spearman correlation estimates when using the Gaussian copula. For ease of exposition, only the correlations for the top 27 websites are presented, which are sites that were visited by at least 5% of panelists in September 2002. Positive pairwise dependence is strong for a number of websites. For example, msn.com, msnbc.com and hotmail.com all have Spearman correlations in excess of 0.6, which is likely because they are either fully or partially owned components of the inter-linked Microsoft Network.

To benchmark the performance of the copula model we extend the validation study found in Danaher (2007) to include the Gaussian copula model. In this study, advertising schedules of between 2 and 15 websites were considered. For each schedule size, 200 different schedules were randomly selected from the 45 sites considered here and different models fit using the September 2002 data to each schedule. These include the Sarmanov with NBD margins and a model assuming independence with NBD margins, as well as our Gaussian copula model (other models reported by Danaher (2007) are not considered here as they are all inferior to the Sarmanov model). Forecasts were made for the November data, adjusting the margins for the change in the mean number of page impressions between the two months, but keeping the estimated dependency structure constant between sites. The performance of the method was judged by computing the RER and EPOR for the estimated total exposure distribution

across the websites in each schedule. Table 5 reports summaries of these metrics in the same manner as Danaher's (2007) Table 4. Again, the Gaussian copula substantially outperforms all the other methods in terms of both estimating reach (lower mean RER values) and the overall exposure distribution (lower mean EPOR values). Moreover, these metrics have lower standard deviations, suggesting that the improvement is consistent.

Last, we note that the effective limit of the number of sites that can be handled by even the approximate Sarmanov model is around 15. In comparison, the Gaussian copula model was successfully applied to model all 45 websites jointly - a massive improvement. Table 6 gives the observed and estimated exposure distribution for this 45 website schedule. As for Table 5, we additionally include the independence model. It can be seen that the Gaussian copula model strongly outperforms the independence model for reach and the full exposure distribution. Moreover, the RER and $EPOR$ values are similar in size to those for the much smaller schedules in Table 5, indicating that the copula model still retains its accuracy in high dimensions.

Finally, two points are worth noting. First, to be computationally feasible, only elliptical copulas can be used for total exposure distributions for advertising campaigns featuring banners on 15 to 45 websites. Second, in the validation study, the copula model need only be fit once and a single Monte Carlo sample of exposures for all 45 sites generated. Monte Carlo iterates of total exposure for any given subset of these 45 websites can then be computed by simply calculating the sum of exposures over only those websites in the smaller schedule. In comparison, the other models have to be re-estimated for each of the 2800 schedules in the validation study. This has huge computational advantages for copula models, for example, when looking for the advertising schedule which optimizes reach.

5 Conclusion

The growing availability of large customer databases with many measured variables per customer has opened up the possibility for using more than just elementary marketing analytics (Davenport and Harris 2007; Rossi and Allenby 2000). When modeling such multivariate data, the dependencies can be both subtle and challenging; failure to adequately capture these can result in poor model fits and forecasts (Park and Fader 2004; Danaher 2007). This is demonstrated by the successful application of many known multivariate distributions from the statistics literature. However, these previous models are limited in that their marginals all have fixed distributions, usually of the same type. Moreover, a large number of marketing applications require discrete, rather than continuous, distributions with complex dependency structures in high dimensions - something that is beyond the scope of traditional models to capture.

Copula models offer a versatile solution to these multiple demands, but, until now, there have been two barriers to using copulas in marketing. These are the multivariate discrete nature and the high dimensionality of much marketing data. While copula models can be easily defined for these cases, it is not possible to reliably estimate them using maximum likelihood techniques. However, in recent research, Pitt et al. (2006) show how Bayesian MCMC estimation can be used to estimate elliptical copula models when the data are discrete. Even so, their approach involves a complex prior on the dependency structure which is not easily applied to higher dimensions. Hence, we develop a fast and efficient alternative MCMC algorithm that enables estimation of multivariate discrete distributions in large dimensions, up to 45 in one application. Therefore, with the developments in this study, the method of copula modeling is a technique whose time has come for marketing.

In summary, the key advantages of copula modeling for marketers are (i) the ability to combine completely different univariate distributions into a well-defined multivariate model, (ii) no approximations are required in high dimensions, and (iii) when using elliptical copulas,

the number of parameters and computational demands do not "blow up" in high dimensions, as can happen for other multivariate probability models used in marketing applications.

The examples used in this study show that copula modeling has merit. In the case of website visit duration and amount purchased, a naive correlation analysis leads to a conclusion that there is no relationship between website "stickiness" and downstream purchase amount. However, a more careful analysis with copula modeling gives online retailers some assurance that their efforts at retaining visitors to their site can have financial benefit. The application of copula modeling to high-dimensional discrete distributions is particularly noteworthy, where our model is substantially more accurate than competing models at estimating multivariate magazine exposure distributions and page views to numerous websites.

Even though we have illustrated the key advantages of copula models in marketing, there are other potential applications, which remain for future research. Consider Goodhardt, Ehrenberg and Chatfield's (1984) "Dirichlet" model, which combines a model for category purchases (the NBD) with a model for brand choice, the Dirichlet Multinomial Distribution (DMD). A disadvantage of the Dirichlet is that the model for brand choice cannot accommodate covariates, such as price and promotion. By contrast, the MNL model developed by Guadagni and Little (1983) can accommodate such marketing mix variables. Even though the NBD and MNL do not come from the same distributional family, copula models offer a way to combine them. This results in a Dirichlet-style model where the DMD is replaced with the MNL, allowing for the inclusion of covariates, a modeling possibility which has previously eluded marketing scientists. In the same way, a copula model could combine a diffusion model for a new product category with a model for brand choice within the category. Lastly, another possible application of copula models is to extend the purchase timing work of Chintagunta and Haldar (1998) from two to many product categories.

Technical Appendix

This appendix details how to derive the conditional Bayesian posterior distribution given at Step 1(b) of Algorithm 2, and also how to implement Algorithm 3 when a Gaussian copula is employed.

For Step 1(b) we outline the derivation found in Pitt et al. (2006, pp. 543-544). Following from Bayes theorem,

$$f(x_{ij}^{\star}|\{x^{\star}\backslash x_{ij}^{\star}\},\Gamma,x) \propto f(x|x^{\star})f(x_{ij}^{\star}|\{x^{\star}\backslash x_{ij}^{\star}\},\Gamma)\,. \tag{A1.1}$$

Now, conditional on the latent variables $x^{\star}$, the data x are independently distributed, so that $f(x|x^{\star}) = \prod_{i=1}^{n}\prod_{j=1}^{m} f(x_{ij}|x_{ij}^{\star})$. For this discrete conditional distribution

$$f(x_{ij}|x_{ij}^{\star}) = \Pr(X_{ij} = x_{ij}|x_{ij}^{\star}) = \mathcal{I}(T^{L} < x_{ij}^{\star} \leq T^{U})\,,$$

where the indicator function $\mathcal{I}(A) = 1$ if A is true, and $\mathcal{I}(A) = 0$ if A false. The bounds for $x_{ij}^{\star}$ are given by the inequality at equation (2.11), so that

$$T^{L} = \Phi^{-1}(F_j(x_{ij}-1))\,,\;\; T^{U} = \Phi^{-1}(F_j(x_{ij}))\,.$$

In the case of a Gaussian copula, the latent variables $x_i^{\star}$ are jointly distributed independent $N(0,\Gamma)$, so that the conditional distribution of the jth element $x_{ij}^{\star}$, given the other elements, is $N(\mu_{ij},\tau_{ij}^2)$, where the mean μ_{ij} and τ_{ij}^2 are the usual conditional mean and variance from a normal distribution; see Greene (2003 p.872). Substituting this into equation (A1.1), results in a $N(\mu_{ij},\tau_{ij}^2)$ distribution constained between lower bound T^{L} and upper bound T^{U}.

To implement Algorithm 3, we first note that, via Bayes theorem,

$$\begin{aligned} f(r_{ij}|\{R\backslash r_{ij}\}, x^\star, x) &\propto f(x|x^\star) f(x^\star|R) p(r_{ij}) \\ &\propto |\Gamma^{-1}|^{n/2} \exp\left\{-\frac{1}{2}\mathrm{tr}(\mathcal{S}\Gamma^{-1})\right\}, \end{aligned}$$

where $\mathcal{S} = \sum_{i=1}^{n} x_i^\star x_i^{\star\prime}$. To implement Step 2(a), we first generate a new proposal value, say r_{ij}^{new}, from a $N(r_{ij}^{\text{old}}, 0.01)$ distribution, where r_{ij}^{old} is the previous iterate value of r_{ij}. This is accepted with probability

$$\alpha = \min\left(1, f(r_{ij}^{\text{new}}|\{R\backslash r_{ij}\}, x^\star, x)/f(r_{ij}^{\text{old}}|\{R\backslash r_{ij}\}, x^\star, x)\right).$$

If it is not accepted, the old value r_{ij}^{old} is retained. This is an implementation of the random walk Metropolis-Hastings step that is a popular computational approach in the MCMC literature. Steps 2(b) and (c) are just straightforward computations.

Algorithm 3 repeats this process to generate all of the upper triangular elements of R one at a time. In addition, if the order in which the elements are generated is randomized at each sweep, the MCMC scheme has better mixing properties; something we do in all our empirical work.

Last, we note here that Algorithm 2 can be extended to cope with a t-copula model. An additional technical appendix outlining this extension is available.

References

Ad Age (2006), "FactPack; 4th Annual guide to Advertising and Marketing", http://adage.com/images/random/FactPack06.pdf (accessed on 22 May 2008).

Albert, James and Siddhartha Chib (1993), "Bayesian Analysis of Binary and Polychotomous Response Data", *Journal of the American Statistical Association*, 88, 669-679.

Bernard, J., Robert McCulloch and X. Meng (2000), "Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage", *Statistica Sinica*, 10, 1281-1311.

Bradlow, Eric T. and Vitala R. Rao (2000), "A Hierarchical Bayes Model For Assortment Choice", *Journal of Marketing Research*, 37, 259-268.

Bucklin. Randolph E. and Catarina Sismeiro (2003), "A Model of Web Site Browsing Behavior Estimated on Clickstream Data", *Journal of Marketing Research*, 40 (August), 249-267.

Chandon, Jean-Louis J. (1986), *A Comparative Study of Media Exposure Models*, New York, NY: Garland.

Cherubini, U., Luciano, E., and Vecchiato, W. (2004), *Copula methods in finance*, New York, NY: Wiley.

Chintagunta, Pradeep K and Sudeep Haldar (1998),"Investigating Purchase Timing Behavior in Two Related Product Categories", *Journal of Marketing Research*, 35 (February), 43-53.

Danaher, Peter J. (1991), "A Canonical Expansion Model for Multivariate Media Exposure Distributions: a Generalization of the Duplication of Viewing Law", *Journal of Marketing Research*, (August), 28, 3, 361-367.

Danaher, Peter J. (1992), "Some Statistical Modeling Problems in the Advertising Industry", *The American Statistician*, 46, 4, 254-260.

Danaher, Peter J. (2002) "Optimal Pricing of Subscription Services: Analysis of a Market Experiment", *Marketing Science*, 21, 2, (Spring), 119-138.

Danaher, Peter J. (2007) "Modeling Page Views Across Multiple Websites With An Application to Internet Reach and Frequency Prediction", *Marketing Science*, 26, 3 (May/June), 422-437.

Danaher Peter J. and Bruce, G.S. Hardie (2005), "Bacon with Your Eggs? Applications of a New Bivariate Beta-Binomial Distribution", *The American Statistician*, 59, 4, November, 282-286.

Danaher, Peter J., Guy Mullarkey and Skander Essegaier (2006), "Factors Affecting Website Visit Duration: A Cross-Domain Analysis", *Journal of Marketing Research*, 43 (May), 182-194.

Davenport, Thomas H. and Jeanne G. Harris (2007), *Competing on Analytics: The New Science of Winning.* Boston, MA: Harvard Business School Press.

Daul, S. De Giorgi, E., Lindskog, F., and McNeil, A. J. (2003), "The grouped t-copula with an application to credit risk", *Risk*, 16, 73-76.

Edwards, Y. and Greg Allenby (2003), "Multivariate Analysis of Multiple Response Data", *Journal of Marketing Research*, 40, 321-334.

Ehrenberg, Andrew S. C. (1988), *Repeat Buying: Facts, Theory and Applications.* (2nd ed.). London: Charles Griffin and Company Limited.

Fader, Peter S and Bruce G.S. Hardie (2007), "How to project customer retention", *Journal of Interactive Marketing*,21, 1, 76-90.

Fang, Hong-Bin, Kai-Tai Fang and Samuel Kotz (2002), "The Meta-elliptical Distributions with Given Marginals", *Journal of Multivariate Analysis*, 82, 1-16.

Frees, E.W. and E.A. Valdez (1998), "Understanding Relationships Using Copulas", *North American Actuarial Journal*, 2, 1, 125.

Genest, Christian and Jock MacKay (1986), "The Joy of Copulas: Bivariate Distributions with Uniform Marginals", *The American Statistician*, 40, 4, 280-283.

Genz, Alan (1992) "Numerical Computation of Multivariate Normal Probabilities", *Journal of Computational and Graphical Statistics*, 1, 141-149.

Genz, Alan and Frank Bretz (2002) "Methods for the Computation of Multivariate t-Probabilities", *Journal of Computational and Graphical Statistics*, 11, 950-971.

Greene, William H. (2003), *Econometric Analysis.* (International 5th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Goodhardt, Gerald J., Andrew S. C. Ehrenberg and Christopher Chatfield (1984), "The Dirichlet: A Comprehensive Model of Buying Behavior." *Journal of the Royal Statistical Society A*, 147 (5), 621-655

Guadagni, Peter M. and John D.C. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data", *Marketing Science*, 2, 3 (Summer), 203-238.

Hong, Yongmiao, Jun Tu and Guofu Zhou (2007), "Asymmetries in Stock Returns: Statistical Tests and Economic Evaluation", *The Review of Financial Studies*, 20, 5, 1547-1581.

Huang, Chun-Yao and Chen-Shun Lin (2006), "Modeling the Audiences Banner Ad Exposure for Internet Advertising Planning", *Journal of Advertising*, 35, 2, 23-37.

IAB (2007) "IAB Internet Advertising Revenue Report", 23 May 2007, http://

www.iab.net/resources/adrevenue/pdf/ IAB_PwC_2006_Final.pdf

Jain, Dipak C. and Naufel J. Vilcassim (1991), "Investigating Household Purchase Timing Decisions: A Conditional Hazard Function Approach", *Marketing Science*, 10 (1), Winter, 1-23.

Jeuland, Abel P, Frank Bass and Gordon Wright (1980), "A Multibrand Stochastic Model Compounding Heterogenous Erland Timing and Multinomial Choice Processes", *Operations Research*, 28, 2, 255-277.

Joe, Harry (1997), *Multivariate Models and Dependence Concepts*, Chapman and Hall.

Johnson, Norman, Samuel Kotz and N. Balakrishnan (1995), *Continuous Univariate Distributions*, vol. 1, 2nd ed., New York, NY: John Wiley.

Johnson, Norman, Samuel Kotz and N. Balakrishnan (1997), *Discrete Multivariate Distributions*, New York, NY: John Wiley.

Kotz, Samuel, Norman Johnson and N. Balakrishnan (2000), *Continuous Multivariate Distributions*, vol., 2nd ed., New York, NY: John Wiley.

Leckenby, John D. and Shizue Kishi (1984), "The Dirichlet-Multinomial Distribution as a Magazine Exposure Model", *Journal of Marketing Research*, 21, 100-106.

Lee, Mei-Ling T (1996), "Properties and Applications of the Sarmanov Family of Bivariate Distributions", *Communications in Statistics: Theory and Methods*, 25, 6, 1207-1222.

Leeflang, Peter S.H., Dick R. Wittink, Michel Wedel and Philippe A. Naert (2000), *Building Models for Marketing Decisions*, Boston, MA: Kluwer Academic Publishers

Lilien, Gary, Philip Kotler and K. Sridhar Moorthy (1992), *Marketing Models*, Englewood Cliffs, NJ: Prentice-Hall.

Moe, Wendy W. and Peter S. Fader (2004), "Dynamic Conversion Behavior at e-Commerce Sites", *Management Science*, 50, 3, 326-335.

Morrison, Donald G. (1979), "Purchase Intentions and Purchasing Behavior", *Journal of Marketing*, 43 (Spring), 65-74.

Morrison, Donald G. and David C. Schmittlein (1987), "Generalizing the NBD Model for Customer Purchases: What Are the Implications and Was It Worth the Effort?", *Journal of Business and Economic Statistics*, 6, 2, 145-159.

Nelsen, R., (2006), *An Introduction to Copulas*, 2nd ed., New York: NY: Springer.

Panagiotelis, Anastasios and Michael Smith (2008), "Bayesian identification, selection and estimation of semiparametric functions in high-dimensional additive models", *Journal of Econometrics*, 143, 291-316.

Park, Young-Hoon and Peter S. Fader (2004), "Modeling Browsing Behavior at Multiple Websites", *Marketing Science*, 23, 3 (Summer), 280-303.

Pitt, Michael, David Chan and Robert Kohn (2006), "Efficient Bayesian Inference for Gaussian Copula Regression Models", *Biometrika*, 93, 3, 537-554.

Poon, Ser-Huang, Michael Rockinger and Jonathan Tawn (2004), "Extreme Value Dependence in Financial Markets: Diagnostics, Models and Financial Implications", *The Review of Financial Studies*, 17, 2, 581-610.

Robert, Christian R. and George Casella (2004), *Monte Carlo Statistical Methods*, (2nd ed.), New York, NY: Springer.

Rossi, Peter E. and Greg M. Allenby (2003), "Bayesian Statistics and Marketing", *Marketing Science*, 22, 3 (Summer), 303-328.

Rossi, Peter E., Greg M. Allenby and Rob McCulloch (2000), "Statistics and Marketing", *Journal of the American Statistical Association*, 95, (June), 635-638.

Rossi, Peter E., Zvi Gilula and Greg M. Allenby (2001), "Overcoming Scale Usage Heterogeneity: A Bayesian Hierarchical Approach", *Journal of the American Statistical Association*, 96, (March), 20-31.

Rust, Roland T. (1986), *Advertising Media Models: A Practical Guide.* Lexington, MA: Lexington Books.

Sarmanov, O. V. (1966), "Generalized Normal Correlations and Two-Dimensional Frechet Classes", *Doklady (Soviet Mathematics)*, 168, 596-599.

Schmittlein, David C., Donald G. Morrison and Richard Colombo (1987), "Counting Your Customers: Who Are They and What Will They Do Next?", *Management Science*, 33, 1, 1-24.

Schweidel David, Peter S. Fader and Eric T. Bradlow (2008a), "Understanding Service Retention Within and Across Cohorts Using Limited Information", *Journal of Marketing*, 71, 1 (January), 82-94.

Schweidel David, Peter S. Fader and Eric T. Bradlow (2008b), "A Bivariate Timing Model of Customer Acquisition and Retention", *Marketing Science*, 27, 5, 829-843.

Sklar, A. (1959), "Fonctions de rpartition n dimensions et leurs marges", *Publications de l'Institut de Statistique de L'Universit de Paris*, 8, 229-231.

Smith, Michael, Sharat Mathur and Robert Kohn (2000), "Bayesian semiparametric regression: an exposition and application to print advertising data", *Journal of Business Research*, 49, 3, 229-244.

Smith, Michael and Robert Kohn (2002), "Parsimonious covariance matrix estimation for longitudinal data", *Journal of the American Statistical Association*, 97, 1141-1153.

Song, Peter Xue-Kun (2000), "Multivariate Dispersion Models Generated from Gaussian Copula", *Scandinavian Journal of Statistics*, 27, 305-320.

Trivedi, P. and Zimmer, D. (2005), "Copula Modeling: An Introduction for Practitioners", *Foundations and Trends in Econometrics*, vol 1, 1-111.

	Bacon				
Eggs	*0*	*1*	*2*	*3*	*4*
Observed Frequencies					
0	254	34	8	0	1
1	115	29	8	0	1
2	42	16	3	4	1
3	13	6	3	1	0
4	6	1	1	1	0
Sarmanov Model Estimates					
0	251.5	37.8	8.7	1.5	0.1
1	112.4	23.6	8.1	2.5	0.5
2	47.2	13.4	5.6	2.0	0.4
3	16.1	6.0	2.8	1.1	0.3
4	3.4	1.7	0.9	0.3	0.1
Gaussian Copula Model Estimates					
0	253.7	34.0	8.4	2.1	0.2
1	111.2	25.0	8.5	2.4	0.4
2	47.3	14.4	5.2	1.9	0.3
3	15.9	6.2	2.8	1.1	0.2
4	3.7	1.6	0.9	0.4	0.1

Table 1: Observed and Fitted Frequencies for Sarmanov Model and Gaussian copula with beta-binomial margins for the bacon and eggs data.

x_1	$F_1(x_1)$	$\Phi^{-1}(F_1(x_1))$	x_2	$F_2(x_2)$	$\Phi^{-1}(F_2(x_2))$
0	0.5448	0.1126	0	0.7857	0.7917
1	0.8139	0.8924	1	0.9363	1.5243
2	0.9398	1.5532	2	0.9837	2.1382
3	0.9882	2.2638	3	0.9974	2.7974

Table 2: Bounds for the latent variables U_j and $X_j^\star$ for the discrete-valued bacon and eggs data. Here, the distribution functions F_1 and F_2 are conditional on the estimated marginal parameter values $\hat{\theta}_1$ and $\hat{\theta}_2$.

Type of Estimate	Correlation
Pearson - empirical	0.08
Spearman - empirical	0.26
Parmetric Spearman with Gaussian copula	0.26
Parmetric Spearman with t_5-copula	0.26

Table 3: Comparison of the empirical Pearson and Spearman correlations with parametric Spearman correlations based on Gaussian and t_5 copulas with log-normal and GEV margins.

	Full Year				*Half Year - Validation Test*			
Exposure Count	Observed	Sarmanov Model	Independent Margins	Gaussian Copula	Observed	Sarmanov Model	Independent Margins	Gaussian Copula
0	*22.1*	16.1	7.4	19.8	*22.2*	15.7	7.2	19.5
1	*6.5*	9.1	7.0	8.6	*6.8*	9.0	6.9	8.6
2	*6.8*	6.9	7.4	7.2	*6.8*	6.8	7.3	7.0
3	*4.5*	6.7	8.6	6.9	*4.6*	6.7	8.5	6.8
4	*13.5*	11.1	13.2	9.7	*13.5*	11.0	13.2	9.3
5	*5.9*	6.9	10.4	6.6	*5.9*	7.0	10.4	6.5
6	*5.6*	5.9	9.2	5.5	*5.6*	5.9	9.2	5.6
7	*4.3*	5.7	8.5	5.0	*4.3*	5.8	8.5	5.0
8	*6.7*	6.0	8.2	4.6	*6.6*	6.1	8.3	4.8
9	*3.9*	5.0	5.7	3.7	*3.7*	5.0	5.7	4.0
10 - 20	*18.1*	20.4	14.1	20.0	*18.0*	20.6	14.5	20.2
RER		0.0770	0.1887	0.0297		0.0835	0.1928	0.0343
EPOR		0.3312	0.5841	0.2542		0.3368	0.5874	0.2566

Table 4: Exposure distributions (in percent) for the 10 magazine example. The left hand side contains results for the fit to the full year data, while the right hand side contains results for the fit to the first half year of data only.

Number of Websites	Model	RER, %	EPOR, %
2-8 Sites	Sarmanov	5.9	19.7
(1400 Schedules)		(0.11)	(0.23)
	Independent	9.6	23.4
		(0.16)	(0.26)
	Gaussian Copula	5.9	17.4
		(0.07)	(0.13)
9-15 Sites	Sarmanov	3.2	11.3
(1400 Schedules)		(0.07)	(0.01)
	Independent	20.6	38.0
		(0.05)	(0.10)
	Gaussian Copula	2.0	8.7
		(0.02)	(0.03)

Table 5: Average values of RER and EPOR for alternative models for the website page view data. Standard deviations of the RER and EPOR values are given in parentheses.

Page Views	Observed Distribution, % Validation	Model, % Gaussian Copula	Model, % Independent
0	14.4	11.2	0.3
1	3.1	4.1	0.5
2	1.9	2.9	0.6
3	1.5	2.3	0.7
4	1.6	2.0	0.8
5	1.3	1.7	0.8
6	1.3	1.6	0.8
7	1.1	1.4	0.9
8	0.9	1.3	0.9
9	1.1	1.2	0.9
10-15	5.0	6.0	5.5
16-20	3.7	3.9	4.4
21-30	5.6	6.2	8.3
31+	57.5	54.2	74.7
RER,%	–	3.8	16.4
EPOR, %	–	10.7	25.9

Table 6: Comparison of observed and estimated exposure distributions for a schedule containing 45 websites.

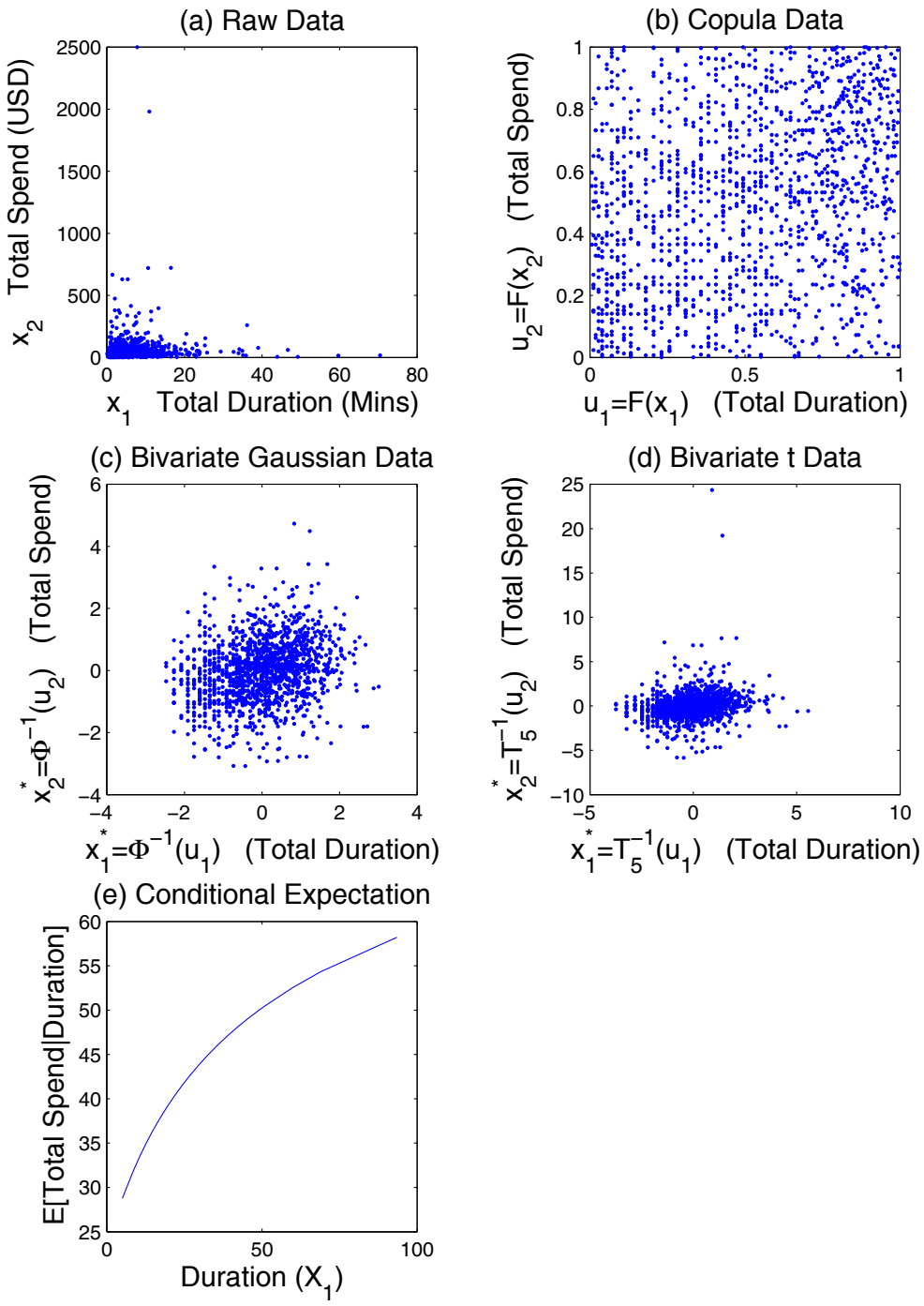

Figure 1: Plot containing results for the bivariate website visit and spend data of motivating example 1. Panel (a) The total duration (X_1) and total spend (X_2) of purchasers at `amazon.com`; Panel (b) The copula data obtained using the probability integral transformation on both margins; Panel (c) The copula data transformed to $\mathcal{R}^2$ using the inverse standard normal distribution function on both margins; Panel (d) The copula data transformed to $\mathcal{R}^2$ using the inverse student t_5 distribution function on both margins; Panel (e) The conditional expectation $E[X_2|X_1]$ for values of duration (X_1) that vary from the lower to upper 5^{th} percentile of duration observed in our data.

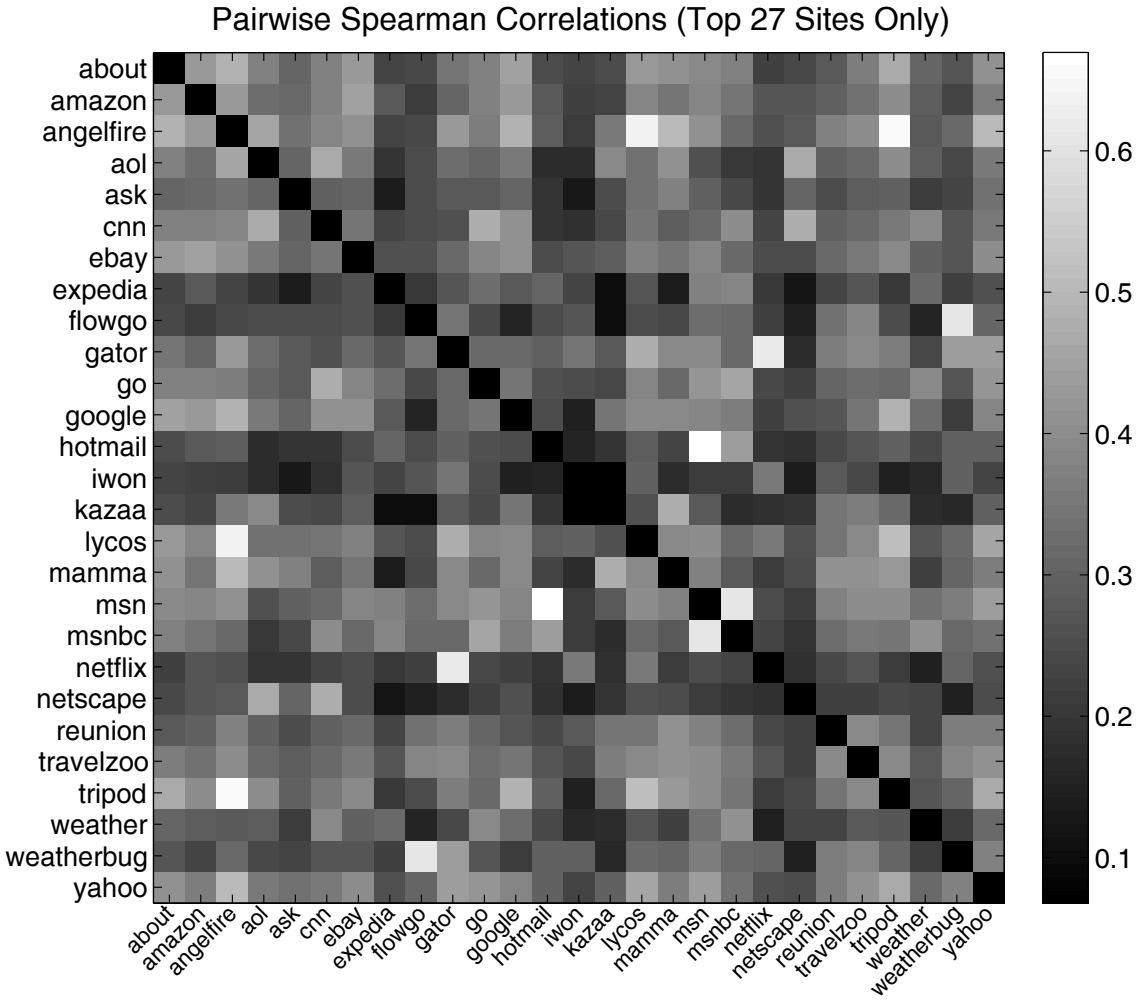

Figure 2: Estimated pairwise Spearman correlations for the top 27 websites using the Gaussian copula model with NBD margins.