Copula Models for Spatial Point Patterns and Processes

Todd H. Kuethe* Todd Hubbs Brigitte Waldorf

Current Draft: May 11, 2009 ©

Abstract

This paper presents a new methodological approach for modeling continuous point-generating processes leading to spatial point patterns. It explores the theoretical foundation and potential applications for copula models in spatial sciences. A copula is a function that combines univariate distributions to obtain a joint distribution with a particular dependence structure. The method is flexible because it separates the choice of dependence among variables from the choice of the marginal distributions of each variable. In addition, copulas are powerful because they are able to capture dependence structures in extreme distributions and in the tails of a distribution. Copula techniques are well established in both financial econometrics and actuarial science, yet the potential of copulas in the context of spatial sciences is relatively unexplored. The study includes an application of spatial copulas to model housing values in an urban area, using complex components such as distance decay, directionality, and edge effects.

Keywords: copula methods, spatial analysis, joint dependence

JEL Codes: C31, R12

*Corresponding author: tkuethe@purdue.edu – The authors wish to thank Raymond Florax for helpful comments on a previous draft of this paper. All remaining errors are our own.

1 Introduction

This paper presents a new methodological approach for modeling continuous point-generating processes leading to spatial point patterns. There are a number of spatial process models developed in recent years to address the effects of spatial dependence in regression analysis, as well as, models developed to interpolate and predict outcomes in space (see Anselin, 2001). These methods are predominately limited to modeling univariate or multivariate distributions in a conditional probability framework. That is, they model the data generating process of random variables given the occurrence of a set of explanatory variables. The focus of this study however is modeling the data generating process of multiple jointly determined variables in space, free of conditional probability assumptions.

Multivariate models are often difficult to estimate in applied work due to complex nonlinearities and nonnormal behavior in observed data, and the modeling of spatial processes is no exception. We attempt to show that researchers can easily draw from current tools for modeling conditional spatial processes to model joint distributions across a geographic space through copula methods. Copula methods provide a set of tools for modeling joint distributions which require only a limited set of information related to univariate marginal distributions and a measure of dependence between variables.

The word copula is a latin noun which means "a link, tie, or bond"(Sklar, 1973). In mathematics and statistics, copula functions tie or bond multivariate distributions based on known univariate distributions. They can be useful when researchers have little information about the nature of the joint dependence structure between random variables but possess some prior knowledge of the one-dimensional marginal distributions, a common occurrence in most fields of applied research.

Copula methods can be used to study a wide array of applied problems in several ways (see Trivedi and Zimmer, 2007). First, copulas can be used to derive joint distributions given the marginal distributions. This property is particularly popular when variables are nonnormal. Second, copulas can incorporate marginal distributions which are difficult to combine in standard joint distributions, such as the widely used multivariate normal distribution or multivariate Poisson distribution. In a copula framework, marginal distributions can take several forms such as beta, Weibull, gamma, exponential, ARCH and GARCH distributions (Trivedi and Zimmer, 2007). The wide range of available marginal distributions then provides a vast number of possible specifications of a joint distribution. Cressie (1993, Section 8.6) presents a modeling framework for multivariate spatial point processes, including Poisson process, Cox process, Markov process, and interrupted process. However, the choice of a multivariate distribution requires strict assumptions of the marginal distributions. For example, a bivariate Poisson process requires *both* components to follow homogeneous Poisson processes. Further, copula functions are flexible in that the marginal distributions can be estimated with a variety of parametric, nonparametric, and semiparametric methods. Third, copulas estimate nonparametric measures of dependence across the entire distribution of observations as opposed to mean correlation measures. Fourth, copulas can model tail dependence in distributions.

In this study, we attempt to show the advantages of copula based methods in spatial sciences, particularly those related to dependence measures which vary throughout the distribution and the effects of tail dependence. Classical measures of spatial dependence, such as Moran's I or LISA statistics, examine the correlation structure of variables between adjacent observations either globally or locally. These measures are a comparison of means which therefore cannot capture the effects of higher order moments, such as skewness or kurtosis. Copula methods on the other hand examine the dependence between random variables across the entire distribution. The methods therefore model the realization of random variables in space which do not require limiting assumptions of underlying distribution.

Copula methods are now common place in other quantitative fields, such as actuarial science where they are used to analyze problems such as the relationship between two individuals' incidence of disease (Clayton, 1978). In finance, copula methods have been used to study a number of problems, such as relationships among multivariate assets in a Value-at-Risk framework (Ane and Kharoubi, 2003). The techniques have gained such popularity that several literature reviews of copula applications can be found. Bouyé et al. (2007) summarize the use of copula methods in finance, and Quinn (2007) review copula applications in health economics. An excellent introduction to copula applications in economics can be found in Trivedi and Zimmer (2007), and a thorough (yet very technical) presentation of copula mathematics is presented by Nelsen (2006).

Despite their widespread use in other quantitative fields, copula methods have not yet been used extensively in spatial sciences. The potential for copula methods in the spatial domain has previously been suggested in a limited number of other geographic sciences, such as geostatistics and geophysics. Kazianka and Pilz (2008) suggest the use of copula-based models for spatial interpolation in a study of accidental releases of radioactive materials. A similar method was also suggested by Bárdossy and Li (2008). Samaniego and Bardossy (2008) use copulas to extract "snapshots" of spatial landscape variability on Earth and Mars through copula based semivariogram techniques, and Bárdossy (2006) uses locally differenced random variables in a copula function in a study of groundwater quality. Bhat and Sener (2008) show that copula methods can be used to consider locational information of neighboring observations as a way to control for spatial dependence. The authors then apply the method to study of the physical activity of teenagers in San

Francisco, California.

Copulas present a number of opportunities for researchers in the spatial sciences. In addition to classic problems of joint dependence, we intend to demonstrate that copulas provide efficient means to address dependence structures which vary over location. We build on previous models of spatial processes and point toward potential avenues of future research.

The remainder of the paper is organized as follows. Section 2 introduces copula functions and defines a number of important concepts. Section 3 provides two brief applications related to modeling urban spatial structure with copula methods. The applications are designed to provide an illustration of just one of the simple yet informative uses of copula methods in spatial sciences. We provide concluding remarks and suggestions for future research in Section 4.

2 Copula Functions

Copula functions provide powerful methods for modeling multivariate distributions, yet in the following section, we limit our discussion to the case of two random variables.[1] We begin with two random variables X and Y with cumulative distributions $F(x) = P[X \leq x]$ and $G(y) = P[Y \leq y]$, respectively, and a joint distribution $H(x, y) = P[X \leq x, Y \leq y]$ (Nelsen, 2006). Although X and Y are separately identifiable, they are realizations of a data generating process which produces the two variables simultaneously.

The foundation of copula modeling is *Sklar's Theorem*, which states that a bivariate distribution can be expressed as a function of the (separate) marginal distributions F and G (Sklar, 1973). Further, if F and G are continuous, then the copula is unique (Trivedi and Zimmer, 2007). The theorem therefore allows for the construction of previously unknown bivariate distributions using known marginals.

Although Sklar (1973) developed copulas in a general distribution theory, we limit our discussion to the analysis of probability distributions of random variables. This limitation is standard in almost all applied studies that employ copula methods because probability distributions, by definition, adhere to the requirements of both marginal distributions and copulas.[2] First, the function's CDF is confined to the unit cube where each marginal is expressed in the range [0,1]. Second, if the CDF has a second derivative, the derivative with respect to the two margins is greater than or equal to 0. Third, at the minimum value of one of the margins, the copula takes the value of 0 for all possible values of the other margin, i.e., the

[1] The interested reader should refer to Trivedi and Zimmer (2007) for a presentation of copula methods in m-dimensional space. However, the interpretation and inference of copula methods becomes challenging as the number of variables increases.

[2] For a thorough description of the conditions for the elements of a copula, called subcopulas, see Nelsen (2006). The author also provides a discussion of the bounds of the probability space of copula functions, the Fréchet-Hoeffding lower and upper bounds.

function is *grounded*. That is, if the probability of a single outcome is 0, the joint probability of all related outcomes is also 0. Further, the dependence structure between the two random variables must fall within a range of perfect negative dependence to perfect positive dependence. Perfect negative dependence takes the value of -1. An elementary example of *near* perfect negative dependence is a car's fuel efficiency and the amount of money necessary to purchase gas per mile. Perfect positive dependence takes the value of 1, and an elementary example of *near* perfect positive dependence is a person's shoe size and the length of his or her feet.

For a joint distribution function, $H(x, y)$, the copula C takes the form:

$$H(x, y) = C[F(x), G(y)|\theta] \tag{2.1}$$

where θ is a scalar parameter called the *dependence parameter*. In copula functions, the dependence parameter is based on the concept of concordance. Two variables are concordant if large values of one variable are associated with large values of the other variable, and small values of one variable are associated with small values of the other variable (Trivedi and Zimmer, 2007). When $\theta > 0$, X and Y exhibit *positive* dependence, and when $\theta < 0$, X and Y exhibit *negative* dependence. In the case of $\theta = 0$, X and Y are independent. Schweizer and Wolff (1981) show that copula functions are an alternative form of two more popular concordance measures, Spearman's ρ and Kendall's τ.

Once the marginal distributions are specified, the researcher must select a copula function. One of the most popular copula classes in empirical studies is the *elliptical* class which includes the *Gaussian* (normal) copula and the *Student* t-copula. The Gaussian copula simply constructs a bivariate normal distribution using Sklar's Theorem. The class is flexible and is well suited to cases in which the random variables are approximately normal. That is the marginal distributions F and G are approximately normal. The Gaussian copula also allows equal degrees of positive and negative dependence. The t-copula specifies an additional dependence parameter which captures fatness in a distribution's tails. A probability distribution with fatty tails exhibits extremely high levels of kurtosis relative to the normal distribution, which exhibits exceptionally thin tails. Higher kurtosis is observed when variance is due to infrequent extreme deviations, as opposed to frequent modestly-sized deviations.

Another general class of copulas common in the literature is the *Archimedean* copula which offers a variety of dependence structures. This class includes the *product* copula which corresponds to independence of the variables examined. The *Clayton* copula is used in the case of strong left tail dependence. The

Gumbel copula is employed in the case of highly correlated variables at high values but less correlated at low values (strong right tail dependence), and the *Frank* copula is applied when tail dependence is weak. The *Farlie-Gumbel-Morgenstern* (FGM) class can be applied when the dependence structure between two marginals is modest in magnitude. More specialized examples include bivariate Poisson, Tobit, and binomial models, but these methods have been applied in only a limited number of studies and are beyond the scope of this study (Trivedi and Zimmer, 2007).

Selecting the proper copula function and fitting the multivariate distribution relies on a mixture of prior knowledge of the problem at hand and statistical methods. When a large sample of data is available, copulas can be fit through maximum likelihood, generalized method of moments, or Bayesian analysis (Trivedi and Zimmer, 2007, Chapter 4). However a number of studies rely more heavily on heuristic methods, such as plots and density estimation of the individual marginal distributions.

3 Applications of Spatial Copulas

In the following section, we provide two illustrations of the use of copula functions to model spatial point patterns. The examples draw from classic models of urban spatial structure and extend such models to a multivariate structure. Both applications model the underlying joint distribution process leading to the formation spatial point patterns of urban form in relation to the distance to the city center, or central business district (CBD). This structure was outlined in the classical model of urban form, the concentric zone model of Alonso (1964).

The first example assumes the realization of each point, the location of a home, is a function of two random variables – the distance to the city center and the distance to each point's nearest neighbor. The second example also uses the distance to the city center, but the second random variable captures the direction of each point in reference to the city center on a two-dimensional plane. These two random variables are the polar coordinates of each observation. The illustrations are simple in nature but are designed to show the potential for copula methods to address a wide range of problems in spatial analysis. In sum, the examples show how copula methods may provide interesting insight even in the simplest of cases by comparing landscape patterns for two classes of land use, high and low value homes.

3.1 Survival Copulas

We begin by outlining a quantitative method for modeling spatial patterns which draws on a time-series method called survival analysis. *Spatial* survival analysis was formalized by Waldorf (2003) and is outlined in the following sections. Further, the foundation of survival analysis has been expanded to a multivariate framework by a number of researchers in health economics and actuarial science in form of survival copulas. We intend to show that these recent developments can therefore be applied to modeling spatial point patterns through multivariate spatial survival analysis in a way that is consistent with temporal modeling.

The translation of survival copula methods to the spatial domain is presented in Figure 1. The foundation of spatial survival analysis is that that distance is mathematically equivalent to temporal duration, and traditional survival models can therefore be used analyze spatial point patterns as a function of distance (see Waldorf, 2003). Researchers in actuarial science have further expanded survival analysis techniques to study multiple dependent lifespans, such as the lifetime of spouses or twins. Building on Waldorf (2003), we take these developments one step further to show that copula methods can be applied to model mulitple distances which are jointly determined. Thus, *spatial copulas* provide an attractive modeling alternative when one distance is jointly determined with another in a point generating process. Examples may include the distance to multiple environmental features or the distance to multiple urban centers in the case of polycentric cities.[3]

[FIGURE 1 ABOUT HERE]

Spatial duration models have been employed in a diverse, yet limited, set of applications. Carruthers et al. (2008) use spatial duration models to examine the role of nearest neighbors in the formation of an urban landscape. Reader (2000) examines patterns of disease spread. Rogerson et al. (1993) analyze the spatial separation of parents and their adult children, and Esparza and Krmenec (1996) investigate the geographical spread of market areas. The applications below show how these spatial survival methods can be combined with survival copula methods to model joint dependence in spatial processes. There are a number of previous studies which employ survival copulas in jointly determined time series framework.[4] For example, the earliest published study of survival copulas by Hougaard et al. (1992) analyzes joint survival times of Danish twins born between 1881 and 1930. Others have used survival copulas to model diverse subjects such as the lifetimes of spouses (Shemyakin and Youn, 2006; Frees et al., 1996) and the tenure of political leaders (Flores, 2008).

[3] It should be noted that copulas can be used to estimate multivariate distributions of random variables in a number of contexts. Although we focus on the use of spatial copulas in the case of joint survival analysis, more complex models with other random variables observed in space can be constructed. Such models are discussed briefly in the closing section of the paper.

[4] Georges et al. (2001) provide a thorough review of the survival copula literature.

3.1.1 Specifying a Survival Function

Duration models, also called longitudinal, survival, or hazard models, are commonly used in engineering and economics to analyze the lifespan of a person or object, called the survival time, and *spatial duration models* use the distance between two points – a one-dimensional projection of two-dimensional space – to serve as a measure of duration, a mathematical equivalent to temporal duration in traditional survival analysis (Waldorf, 2003). The left hand side variable in a spatial duration model is distance D which captures an observation passing between two states s_1 and s_2. In a point-generating process, a new point can emerge at any distance from the already existing points, and it is a realization of a switch between the two states of "no point generated" and "point generated." The distance between the two states s_1 and s_2 is not constant but varies over observations. Thus, D is a continuous, nonnegative random variable for which outcomes are not given a priori.[5]

Spatial survival analysis specifies the probability that the distance D is less than a value d, and the *survival function* $S(d)$ considers a complement of the distribution function such that

$$S(d) = P(D \geq d) = 1 - F(d) \tag{3.1}$$

The related *hazard function*, $h(d)$, is defined as

$$h(d) = -\frac{\partial \ln S(d)}{\partial d} = \frac{f(d)}{S(d)} \tag{3.2}$$

The hazard function for each d can also include explanatory variables. This is commonly referred to as the *proportional hazard function* given by Cox (1972).

$$h(d, Z) = e^{\beta Z} b(d) \tag{3.3}$$

where $b(d)$ is the so-called "baseline" hazard function and β is a vector of unknown parameters describing the effect of exogenous variables Z. Three types of covariates can be included in Z: attributes which do not vary over distance, distant dependent variables that change during a spell, and distance varying covariates that are explicit functions of distance, $z(d)$.

Survival functions are also quite flexible because they have the ability to handle censored observations (for

[5] In a time series context, duration models are often used to examine the lifespan of a person which consists of the time spent between when a he or she is born, where state s_1 represents being alive, and dies, where s_2 represents death. Spatial duration models follow the same construct, yet substitute distance D for time T as a measure of duration.

a thorough review from a time series perspective, see Wooldridge, 2002). An observations is "left censored" if it is in state s_1 at the beginning of the sample period. An observations is "right censored" if it remains in state s_1 at the end of the observation area D. Spatial duration models are often much simpler to estimate than the time series equivalent because distances are not likely to be left censored. However, when the point patterns are observed in a bounded area, distances may be right censored. In spatial sciences, points near the observation area may be subject to the *edge effect* when a researcher suspects the "natural" economic process extends beyond the geographic bounds of the observation area. Spatial survival models can therefore use methods developed for the time domain to address right censoring as a way to control for edge effects. Such methods are be discussed in greater detail in Section 3.2.1.

3.1.2 Multivariate Survival Functions through Copulas

Nelsen (2006) shows that lifetimes of (two) dependent individuals or objects can be modeled with copula functions based on the *joint survival function*

$$\bar{H}(d_1, d_2) = P[D_1 \geq d_1, D_2 \geq d_2] \tag{3.4}$$

where D_1 and D_2 are durations with (marginal) survival functions $\bar{F}$ and $\bar{G}$, respectively. The construction of $\bar{F}$ and $\bar{G}$ follows Equation (3.1). The marginal functions are then tied with a *survival copula*

$$\begin{aligned} \bar{H}(d_1, d_2) &= 1 - F(d_1) - G(d_2) + H(d_1, d_2) \\ &= \bar{F}(d_1) + \bar{G}(d_2) - 1 + C(F(d_1), G(d_2)) \\ &= \bar{F}(d_1) + \bar{G}(d_2) - 1 + C(1 - \bar{F}(d_1), 1 - \bar{G}(d_2)) \\ &= \hat{C}(\bar{F}(d_1), \bar{G}(d_2)) \end{aligned} \tag{3.5}$$

The estimation of the joint survival function is carried out in two stages. In the first stage, the marginal survival functions are estimated through parametric or nonparametric methods. In the second stage, the estimated survival functions are joined according to a copula function.

3.2 Multivariate Modeling of the Urban Form

The following sections present two applications of spatial copula methods. Both applications model the underlying process of the point generating process leading to an urban landscape. The foundation of the classical model of the urban spatial structure is bid-rent theory, often attributed to the early work von Thünen and later formalized by Alonso (1964). The theory assumes that land use activities have different needs to locate near the center of the city, the central business district (CBD), and economic agents bid for land according to their needs. Each group is assumed to consist of homogeneous agents (commercial and manufacturing firms, as well as homeowners) who are costlessly mobile (Anas et al., 1998). The result is a land use or land rent gradient which declines in a more or less predictable fashion outwards from the CBD. Land is therefore allocated such that each actor maximizes either utility or profit subject to some transportation cost budget constraint.

The applications compare two forms of land use, high and low value homes, based on 1,671 observation of houses sold in 2006 in Tippecanoe County, Indiana. The observations were obtained from the local Multiple Listing Service and span the entire year. The data is shown in Figure 3. In Figure 3a, a dot identifies houses which sold at or above the median value, and in Figure 3b, a dot marks the location for homes which sold below the median price. By dividing the sample price, we are able to compare the spatial structure of high value and low value homes. As previously stated, the ability to compare the data generation process for multiple landuses is one of the many uses of copulas.

The city's center, as defined by the county courthouse, is shown with a star in both figures. We believe the county courthouse provides a good proxy of the CBD. It is the central point of business activity, as well as civic institutions.

[**FIGURE 3 ABOUT HERE**]

3.2.1 Two Measures of Distance and the Urban Form

One of the key insights of the classic model of urban form is that as individuals (firms or households) locate further from the CBD, the amount of land consumed by each increases. Thus, as one moves away from the CBD, each parcel size increases, leading to a greater distance between nearest neighbors. A recent study of this phenomenon which draws from spatial duration methdos is presented by Carruthers et al. (2008). The authors estimate spatial hazard functions for the distance to each point's nearest neighbor while controlling for a number of factors, such as the age of a structure, commuting time, and the distance to the central business district.

Our model, on the other hand, suggests that the distance to each observation's nearest neighbor is not a conditional spatial process, but the realization of two jointly determined processes derived from the distance to the nearest neighbor and the distance to the city center. That is, economic agents select a parcel of land as a function of the distance to the CBD and the distance to their nearest neighbor *simultaneously.* Each point is therefore a realization of two jointly determined processes. We believe this construct may better describe the adaptation of a preexisting built environment which includes a mix of both old and new structures which create an urban landscape.

Figure 2 shows how the two distance measures are calculated for each observation. Take for example observation 1. The first measure is given by d_{c1}, the distance from point 1 to the CBD. The second measure d_{12} gives the distance to observation 1's nearest neighbor – observation 2. Although it can be seen that observation 1's nearest neighbor is observation 2, observation 2's nearest neighbor is observation 3. Thus, the two measures recorded for observation 2 are d_{c2} and d_{23}.

[FIGURE 2 ABOUT HERE]

Functional Form

As previously stated, the copula function models the dependence structure of multiple random variables given their individual marginal distributions. We must therefore fit the two random variables, the distance to the CBD and the distance to the nearest neighbor, to a distribution as part of the first step of the analysis. The marginal distribution of both distance variables are fitted to a Weibull distribution via maximum likelihood. The Weibull distribution is common in survival analysis and offers a number of attractive features outlined below. The probability density function of the Weibull distribution for each distance duration $D = d$ takes the form

$$f(d|k, \lambda) = \begin{cases} \frac{k}{\lambda}(\frac{d}{\lambda})^{k-1}e^{(d/\lambda)^k} & \text{if } d \geq 0 \\ 0 & \text{if } d < 0 \end{cases} \tag{3.6}$$

where $k > 0$ is the *shape parameter* and $\lambda > 0$ is the *scale parameter.* The shape parameter determines the duration dependence of d. When $k < 1$, D exhibits negative duration dependence which implies that the probability a distance continues or increases declines as distance increases. When $k > 1$, D exhibits positive duration dependence, and in the case of $k = 1$, the duration dependence is constant.

The nearest neighbor survival function is adapted to control for potential right censoring related to edge effects. We define a censored observation as any house which is located closer to the edge of the observation area than to its nearest observed neighbor. This augmentation allows for the potential that the censored observation's nearest neighbor is actually located outside of the study area, and thus, the economic process continues beyond the geographic boundary analyzed. The number of censored observations is quite small though, comprising approximately 1% of the low value houses and approximately 2% of high value houses.

The moments of the estimated marginal distributions, along with summary statistics, for both variables is shown in Table 1.

[TABLE 1 ABOUT HERE]

The estimated shape parameters for the two models suggest the distance to the city center exhibits positive duration dependence for both high and low value homes. Thus, as distance increases, the probability that it continues to increases is positive. However, the distance to the nearest neighbor exhibits negative duration dependence in both models. It should be noted that the estimated survival functions fit the data to a Weibull distribution, but the model is not conditional on additional covariates. In a number of cases the researcher may be interested in the effects of additional factors and therefore elect to estimate a Cox proportional hazard model as presented in Equation (3.3). This technique is still possible for copula methods, yet the multivariate distribution function would be interpreted as a joint baseline hazard function.

In order to build a copula function, we need some prior knowledge on the dependence structure between the two distance measures, and we measure correlation with Spearman's ρ reported in Table 1. The measure assesses how well an arbitrary monotonic function could describe the relationship between the two variables, free of assumptions of the underlying distributions.

The popularity of copula methods in other disciplines has lead to the development of several specialized software routines in popular statistical packages such as Stata, Limdep, and SAS (Trivedi and Zimmer, 2007). For our analysis we use the **R** package `copula` which offers a suite of estimation techniques for several classes of copula functions. Yan (2007) provides a thorough description of the package and a number of numerical examples.

Once we have identified the marginal distributions and a measure of dependence between the two variables, we build a t-copula with one dependence parameter ρ and degrees of freedom parameter $\upsilon = 5$. The t-copula is popular because it allows for tail dependency based on the degrees of freedom parameter. For large values of υ, the function converges toward a Gaussian distribution, yet small values of υ increase the tail dependency. In order to allow the dependence parameter to vary over the sample, we impose an unclassified dispersion matrix of ρ. The copula function therefore allows for strong tail dependence in the multivariate distribution, yet the correlation is allowed vary over its range in an unstructured manner. Figure 4 shows two and three dimensional presentations of both copulas.

[FIGURE 4 ABOUT HERE]

The dependence relationship is shown along the vertical axis of the three dimensional graph and by isolines in the two dimensional projection. The copula functions represent a high degree of positive dependence when both distance variables are small (i.e., close to the city center and close to the nearest neighbor) and when both distance variables are large (i.e., far from the city center and far from the nearest neighbor). This relationship can be interpreted as a high degree dependence near the city center and at the urban fringe, but a weak dependence level between the two extremes of the built environment. This construct represents the greatest advantage of copula methods. The researcher can allow the dependence structure to vary over the landscape in a fashion that is consistent with both the theory of urban spatial structure and the observed data.

Results

Once the copula is constructed, we can draw random samples of the joint distribution for at least three distinct types of analysis. First, a *comparison* can be made across alternate copula functions or land uses. In our case, we are interested in comparing the data generation process for low value and high value homes. Alternatively the methods could be used to compare alternative locations or points in time. Second, the random draws can be used for *simulation* purposes in Monte Carlo analysis of new estimation techniques or theoretical models. Third, the draws can be used for *prediction* over space or time. That is, once we have identified the data generating process, we can take marginal draws from the multivariate distribution to predict possible future points in space or unobserved points in the case of a small or limited sample size.

From the joint distribution constructed in the previous section, we draw 1,000 random samples which are summarized in Table 1.[6] The multivariate distribution does a fair job of fitting the moments of the observed data in both models. The multivariate pdfs for the random draws are also shown in Figure 5.

[FIGURE 5 ABOUT HERE]

Figure 5a shows the multivariate density for low value homes, and Figure 5b shows the multivariate density function for high value homes. The multivariate density functions can be interpreted as a sequence of univariate distribution functions for given values of either variable. For example, if one selects a distance of 17,000 feet from the city center, this "slice" of the multivariate pdf represents the univariate pdf for the nearest neighbor distance when the house is 17,000 feet from the city center. The multivariate density functions for both low and high value homes indicate that both low and high value homes are likely located near the city center and clustered near neighboring observations. However, the concentration of higher

[6] A thorough description of copula methods for generating random numbers can be found in (Trivedi and Zimmer, 2007, Appendix A).

value homes appears to be slightly farther from the city center and exhibit greater distance to neighboring observations.

The random draws can also be used to compare the observed survival functions with the multivariate estimates which account for joint dependence. The estimated functions are plotted against the observed survival functions for both models in Figure 6. Visually, both models appear to provide a reasonable fit to the data.

[FIGURE 6 ABOUT HERE]

3.2.2 Modelling Distance and Direction

In this example, we intend to show that it is possible to model other random variables observed in space in addition to two measures of distance. We continue with the previous model of urban form but intend to incorporate both distance and directional components of each observation. The directional aspects of the urban spatial structure in relation to the classical model of urban form were introduced by Hoyt (1939). The directional aspects of the urban form lead to clustering of land use in space which is a function of both distance to the CBD and direction along a transportation corridor.

Using the same data presented in Figure 3, we calculate each observation's polar coordinates. The polar coordinate system is a two-dimensional coordinate system in which each point on a plane is determined by two elements: angle and distance. Shown in Figure 7, each point j is uniquely determined by the distance from the central business district d_{cj} and an angle measured from a positive ray from the x coordinate of the central business district, ϕ_j. In our example, ϕ is determined by angular measure from an eastern ray from the Tippecanoe County Courthouse, and the angular measure is bounded by $0 < \phi \leq 360$. The distance measure is the same one previously described.

[FIGURE 7 ABOUT HERE]

Functional Form

The distance to the CBD for both low and high value homes is modeled using the survival function presented in the previous example. The angle measures, however, are shown to be approximately normal by means of the Kolmogrov-Smirnov test. The marginal distributions are therefore fitted according to the mean and standard deviation of each distribution. The results for low value houses and for high value houses are summarized in Table 2.

[TABLE 2 ABOUT HERE]

In both cases, we again select a t-copula with an unstructured dispersion matrix based on Spearman's ρ. The t-copula's ability to measure tail dependence is particularly attractive for modeling polar coordinates. Because the copula allows for strong and equal levels of tail dependence, the dependence parameter may vary over the distribution but take similar values at the tails of the distribution. This feature then allows similar directions at the extremes of the distribution, say $\phi = 1$ and $\phi = 359$, to exhibit similar dependence structure. The estimated t-copulas for low and high value homes are presented in Figure 8.

[**FIGURE 8 ABOUT HERE**]

Results

Figure 9 shows the estimated density functions for both multivariate distributions based on 1,000 random draws, which are summarized in Table 2.

[**FIGURE 9 ABOUT HERE**]

In the previous example, the estimated density functions were quite similar for both low and high value homes which is consistent with the classical model of urban form. However, in the current example the estimated density functions are much different when comparing low and high value homes, Figures 9a and 9b, respectively. This would suggest homes cluster differently in space based on their value yet share similar dynamics with respect to distance to their neighbors. Clustering is particularly evident when comparing the tail behavior of the bivariate distributions for low and high value homes. For example, the angle measure for low value homes exhibits greater left tail dependence. This indicates clustering of high value homes in the Northeastern region of the observation area. This appears to be consistent with the map shown in Figure 3.

4 Conclusion

In the previous sections, we have outlined a new methodological approach for modeling spatial patterns and point-generating processes leading to spatial point patterns. We presented the theoretical foundation and potential applications for copula models in spatial sciences. Although copula methods are well established in other quantitative fields, the potential of in this context is relatively unexplored. We show that copula methods can be used to garner new information about jointly determined data generation process in the spatial domain.

There are several appealing aspects of copula methods for the spatial sciences. Copula methods provide a flexible and powerful set of tools for modeling jointly determined random variables because the methods separate the choice of dependence from the choice of the marginal distributions of each variable. In addition, copulas are powerful because they are able to capture dependence structures across the entire distribution, such as tail dependence or extreme value dependence.

Our study includes two brief applications of copula methods for the determination of an urban landscape. The examples are simple applications of copula methods in spatial analysis, yet they provide evidence of the power of spatial copula methods. The examples focus on the use copulas to compare the data generating process for two classifications of residential land use: low and high value homes. However, copula methods can be used to make other comparisons in a spatial framework. For example, one can compare the data generating process for two variables, such as residential and commercial parcels, or one process at two locations, such as residential property Chicago and New York. One could also compare a particular land use in the same study area at two points in time, such as commercial land use in Los Angeles in 1950 and 2000.

Copula methods however are not limited to studies which compare landscapes. Copula methods can also be used to generate spatial data for Monte Carlo applications or to predict stocks and flows of random variables in space. For example, the flexibility of copula methods may provide alternative ways to model space-time processes such as migration or disease spread which are examples of spatial flows. A disease spread model, for example, could examine the joint distribution of both spatial spread and temporal duration of a set of outbreaks relative to some initial outbreak in a given location and time.

We believe spatial copula methods can be used to study a number of jointly determined random variables in space beyond the scope of this paper. For example, one can incorporate other models from survival analysis, such as competing risk models for durations which can be terminated in two or more ways or multi-episode models where more than one episode is observed. Examples include polycentric cities (competing risk models) and phenomena which move over time (multi-episode models). In sum, there is a vast potential for copula methods within spatial analysis which has not yet been explored.

References

Alonso, W. (1964). *Location and land use.* Harvard University Press Cambridge.

Anas, A., R. Arnott, and K. Small (1998). Urban spatial structure. *Journal of Economic Literature*, 1426–1464.

Ane, T. and C. Kharoubi (2003). Dependence structure and risk measure. *The Journal of Business 76*(3), 411–438.

Anselin, L. (2001). *A Companion to Theoretical Econometrics*, Chapter Fourteen Spatial Econometrics, pp. 310–330. Blackwell Publishing.

Bárdossy, A. (2006). Copula-based geostatistical models for groundwater quality parameters. *Water Resources Research 42*(11).

Bárdossy, A. and J. Li (2008). Geostatistical interpolation using copulas. *Water Resources Research 44*(7), W07412.

Bhat, C. and I. Sener (2008). A copula-based closed-form binary logit choice model for accommodating spatial correlation across observational units. *Technical paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, June*.

Bouyé, E., V. Durrleman, A. Nikeghbali, G. Riboulet, T. Roncalli, B. Mitterrand, and S. Investments (2007). Copulas for finance: A reading guide and some applications. *Groupe de Recherche Operationnelle, Credit Lyonnais*.

Carruthers, J., S. Lewis, G. Knaap, and R. Renner (2008). Coming undone: A spatial hazard analysis of urban form in american regions. *U.S. Department of Housing and Urban Development Working Paper # REP 08-03*.

Clayton, D. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika 65*(1), 141–151.

Cox, D. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 187–220.

Cressie, N. (1993). *Statistics for Spatial Data.* Wiley-Interscience.

Esparza, A. and A. Krmenec (1996). The spatial extent of producer service markets: hierarchical models of interaction revisited. *Papers in Regional Science 75*(3), 375–395.

Flores, A. (2008). Copula functions and bivariate distributions for survival analysis: An application to political survival.

Frees, E., J. Carriere, and E. Valdez (1996). Annuity valuation with dependent mortality. *Journal of Risk and Insurance*, 229–261.

Georges, P., A. Lamy, E. Nicolas, G. Quibel, T. Roncalli, B. Mitterrand, J. Jouanin, and G. Rapuch (2001). Multivariate survival modelling: a unified approach with copulas.

Hougaard, P., B. Harvald, and N. Holm (1992). Measuring the similarities between the lifetimes of adult Danish twins born between 1881-1930. *Journal of the American Statistical Association*, 17–24.

Hoyt, H. (1939). *The structure and growth of residential neighborhoods in American cities.* Washington Federal Housing Administration.

Kazianka, H. and J. Pilz (2008). Spatial interpolation using copula-based geostatistical models.

Nelsen, R. (2006). *An introduction to copulas.* Springer Science+ Business Media, Inc.

Quinn, C. (2007). The health-economic applications of copulas: Methods in applied econometric research. *HEDG Working Paper 07/22*.

Reader, S. (2000). Using survival analysis to study spatial point patterns in geographical epidemiology. *Social Science & Medicine 50*, 985–1000.

Rogerson, P., R. Weng, and G. Lin (1993). The spatial separation of parents and their adult children. *Annals of the Association of American Geographers*, 656–671.

Samaniego, L. and A. Bardossy (2008). *GI-Days 2008. Proceeding of the 6th Geographic Information Days, Munster Germany*, Chapter Characterizing the spatial variability: Examples from the Earth and Mars, pp. 1–5. Institut for Geoinfromatik.

Schweizer, B. and E. Wolff (1981). On nonparametric measures of dependence for random variables. *The Annals of Statistics*, 879–885.

Shemyakin, A. and H. Youn (2006). Copula models of joint survival analysis. *Applied Stochastic Models in Business and Industry 22*(2), 211–224.

Sklar, A. (1973). Random variables: Joint distribution functions and copulas. *Kybernetika 9*(6), 449–460.

Trivedi, P. and D. Zimmer (2007). *Copula modeling: an introduction for practitioners.* Now Publishers Inc.

Waldorf, B. (2003). Spatial patterns and processes in a longitudinal framework. *International Regional Science Review 26*(3), 269.

Wooldridge, J. (2002). *Econometric analysis of cross section and panel data.* MIT press.

Yan, J. (2007). Enjoy the joy of copulas: with a package copula. *J. Stat. Softw 21*, 1–21.

Table 1: Distance and Distance Application Results

	Low Value		High Value	
	City Center	Nearest Neighbor	City Center	Nearest Neighbor
Observed				
Mean (feet)	17,481.00	565.95	23,405.00	915.35
Std. Dev.	12,276.54	1,335.00	11,823.02	1,968.22
Weibull Parameters				
Shape	1.57	0.83	2.07	0.75
Scale	19,620.03	495.72	26,423.74	738.14
Spearman's ρ	0.16		0.22	
Kendall's τ	0.11		0.14	
Predicted				
Mean (feet)	17,852.20	556.40	23,907.20	854.70
Std. Dev.	11,232.28	664.18	12,359.81	1,168.00

Table 2: Distance and Direction Application Results

	Low Value		High Value	
	City Center	Angle	City Center	Angle
Observed	(feet)	(degrees)	(feet)	(degrees)
Mean	17,481.00	26.34	23,405.00	208.21
Std. Dev.	12,276.54	119.98	11,823.02	92.9
Weibull Parameters				
Shape	1.57		2.07	
Scale	19,620.03		26,423.74	
Spearman's ρ	0.00		–0.11	
Kendall's τ	–0.02		–0.07	
Predicted	(feet)	(degrees)	(feet)	(degrees)
Mean	17,660.35	25.91	23,613.90	211.56
Std. Dev.	11,644.11	122.09	11,910.72	88.68

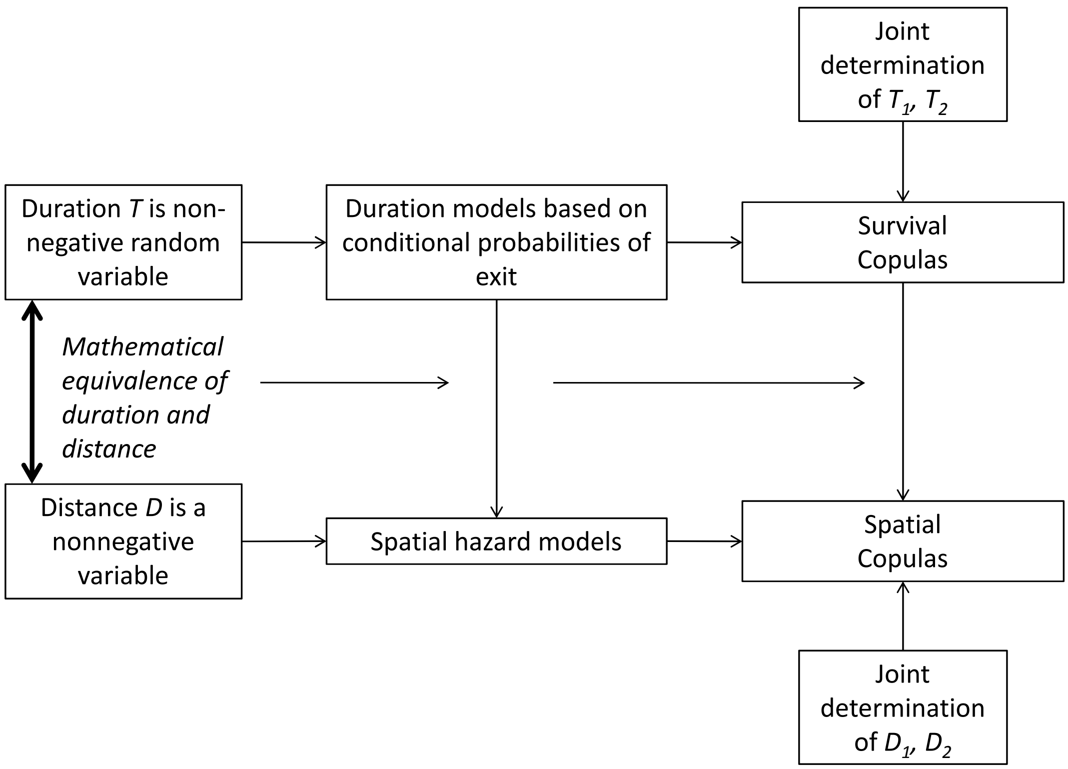


Figure 1: Transfer of Survival Copulas from a Temporal to a Spatial Setting

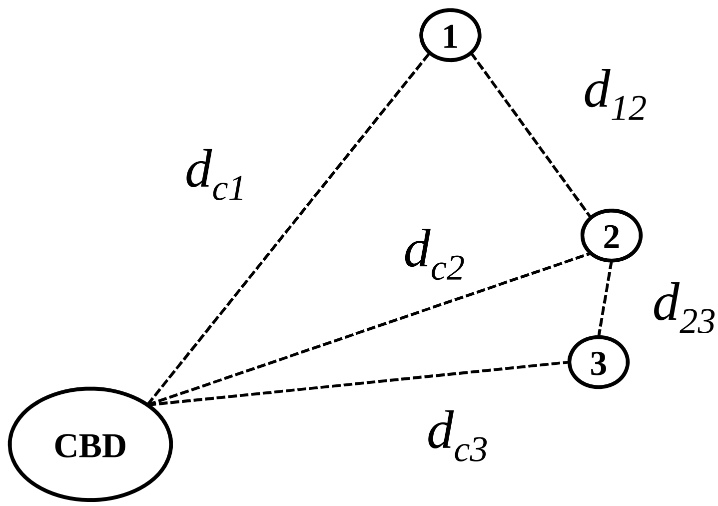
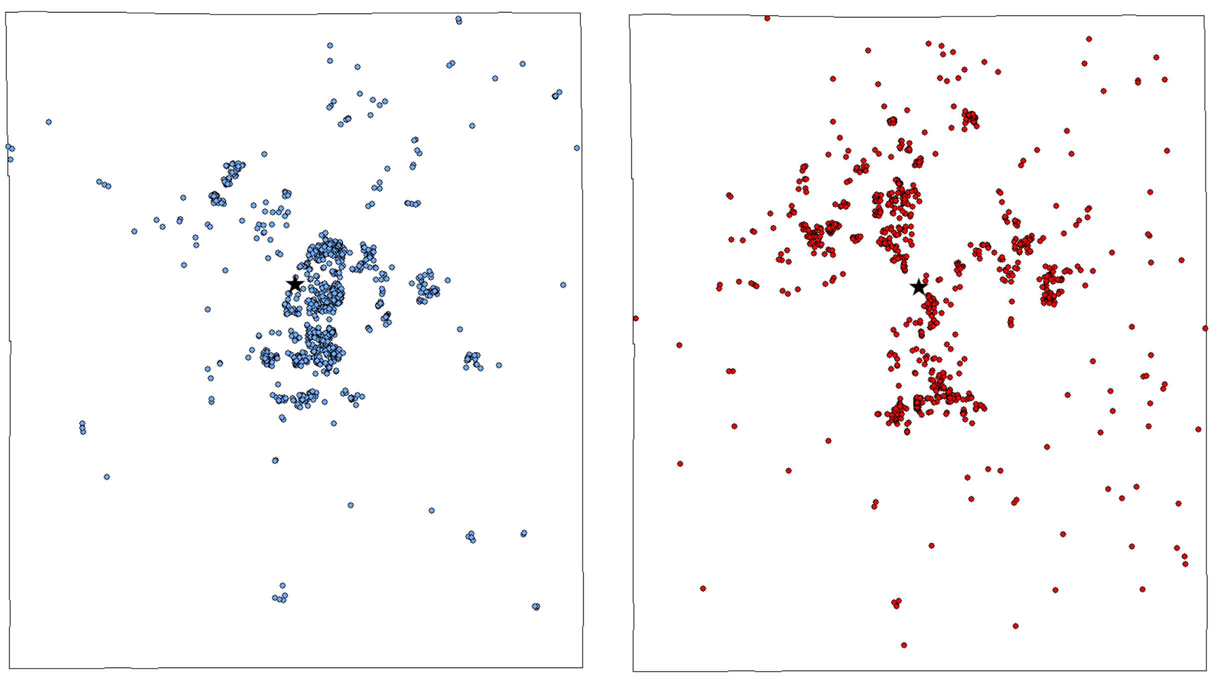


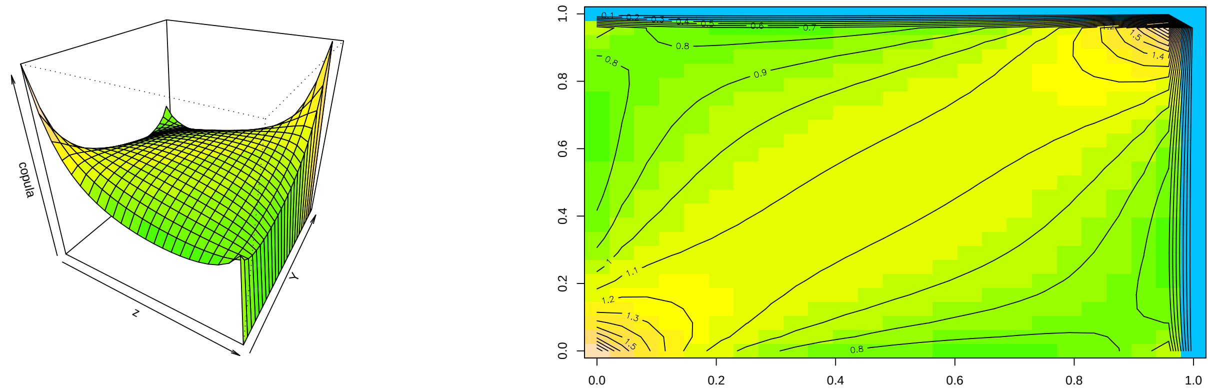
Figure 2: Calculations of Distance Durations



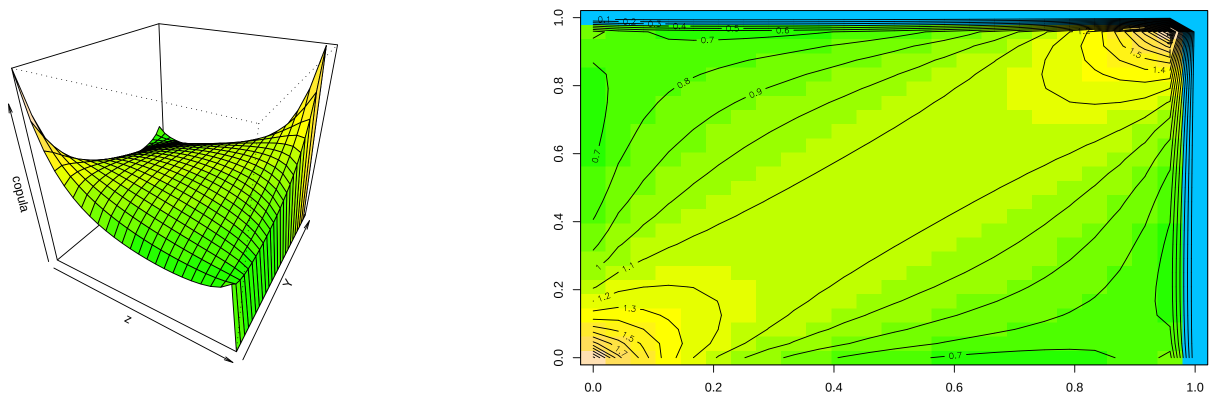
(a) Low Value

(b) High Value

Figure 3: Houses for Sale and City Center, Tippecanoe County, Indiana 2006

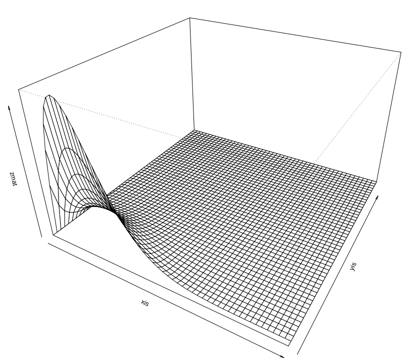


(a) Low Value

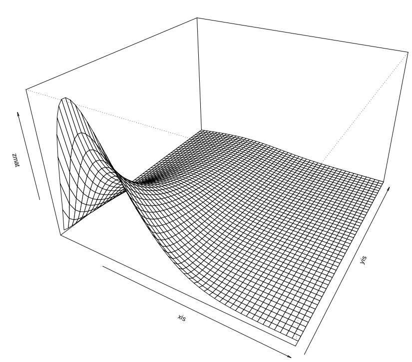


(b) High Value

Figure 4: t-Copulas, $\rho_1 = 0.16$, $\rho_2 = 0.22$, $\upsilon_1 = \upsilon_2 = 5$

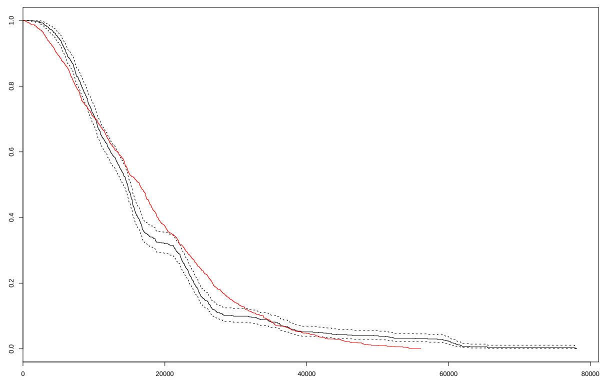


(a) Low Value

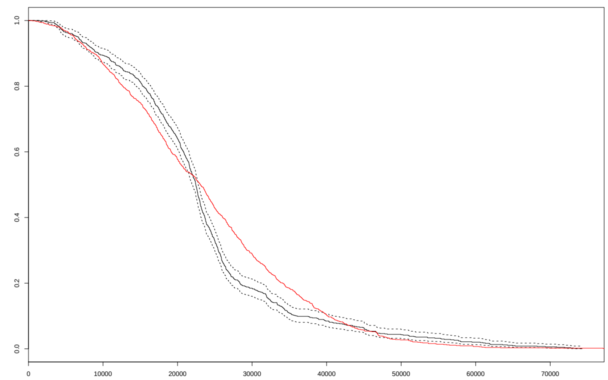


(b) High Value

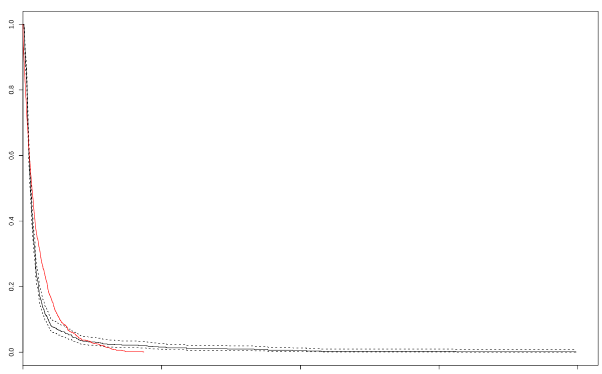
Figure 5: Estimated Density Function of Multivariate Distributions



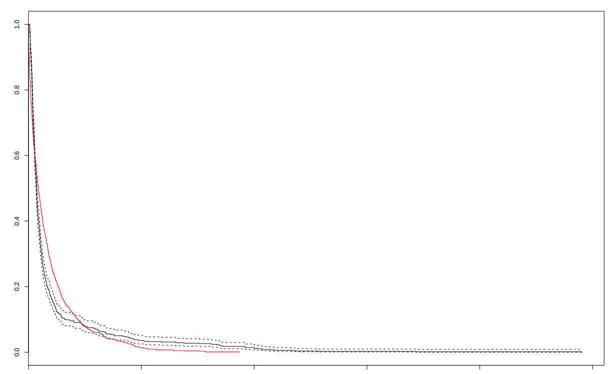
(a) Low Value – City Center



(b) High Value – City Center



(c) Low Value – Nearest Neighbor



(d) High Value – Nearest Neighbor

Figure 6: Simulated (red) and Observed (black) Survival Functions

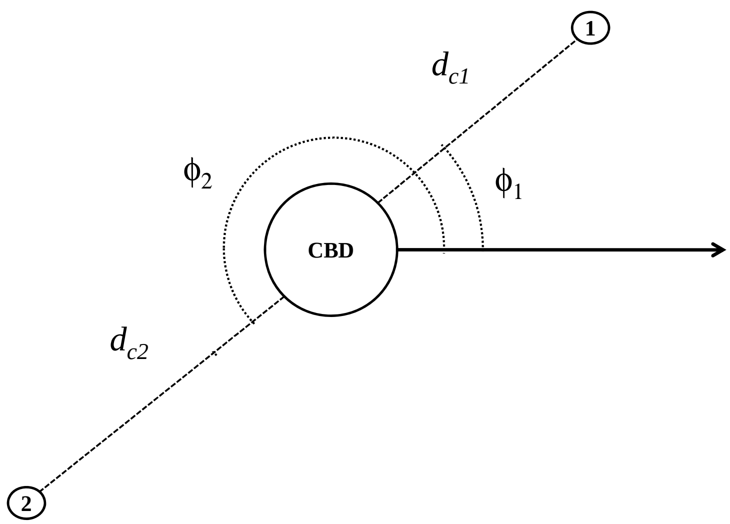
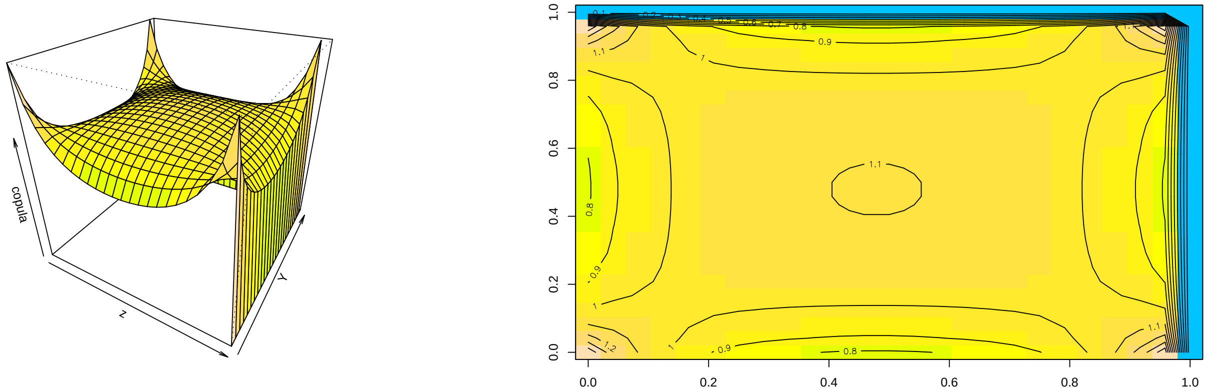
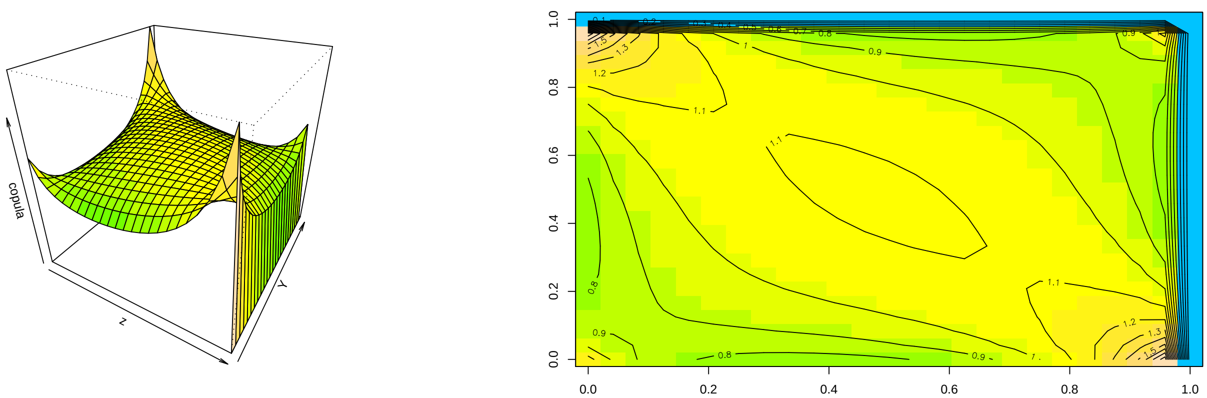


Figure 7: Calculation of Polar Coordinates

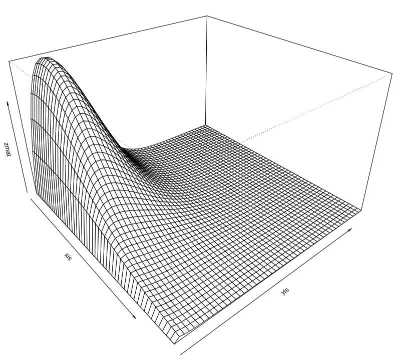


(a) Low Value

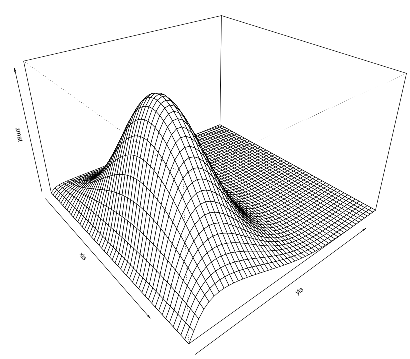


(b) High Value

Figure 8: t-Copulas, $\rho_1 = 0.00$, $\rho_2 =$–0.11, $\upsilon_1 = \upsilon_2 = 5$



(a) Low Value



(b) High Value

Figure 9: Estimated Density Function of Multivariate Distributions