Copula Regression with Discrete Outcomes

by

Lu Yang

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Statistics)

at the

UNIVERSITY OF WISCONSIN–MADISON

2017

Date of final oral examination: 06/06/2017

The dissertation is approved by the following members of the Final Oral Committee:

      Edward W. Frees, Professor, Risk and Insurance

      Zhengjun Zhang, Professor, Statistics

      Peng Shi, Associate Professor, Risk and Insurance

      Chunming Zhang, Professor, Statistics

      Jee-Seon Kim, Professor, Educational Psychology

Abstract

Multivariate discrete outcomes are common in a wide range of areas including insurance, finance, and biology. When the interplay between outcomes is significant, quantifying dependencies among interrelated variables is of great importance. Due to their ability to accommodate dependence flexibly, copulas are being applied increasingly.

Yet the application of copulas on discrete data is still in its infancy; one of the biggest barriers is the identifiability of copulas, calling into question model interpretations and predictions. In this dissertation, we study the issue of identifiability in a regression context. As the marginal distributions vary with covariates, inclusion of continuous regressors provides a region of support for copula identifiability. We establish conditions under which the copula regression model is identifiable for discrete outcomes.

In addition, since the properties of continuous outcomes do not carry over to discrete outcomes, specification of a copula model has been a problem. We propose a nonparametric estimator of copulas to identify the "hidden" dependence structure for discrete outcomes and develop its asymptotic properties. The proposed nonparametric estimator can also serve as a diagnostic tool for selecting a parametric form for copulas. Analysts can check adequacy of fit by comparing the fitted parametric copula with the nonparametric estimator. The comparison suggests improvements for the fitted model.

In the simulation study, we explore the performance of the proposed estimator under different scenarios and provide guidance on when the choice of copulas is important. The performance of the estimator improves as discreteness diminishes. A practical bandwidth selector is also proposed.

An empirical analysis examines a dataset from the Local Government Property Insurance Fund (LGPIF) in the state of Wisconsin. The LGPIF offers different types of coverage for local government properties. We apply the nonparametric estimator to model the dependence between claim frequencies from different types of insurance coverage.

Contents

List of Tables

List of Figures

Acknowledgments

My last five years at the University of Wisconsin-Madison have been very rewarding. This accomplishment would not have been possible without the help of many people.

First, I want to thank my advisors, Professors Edward Frees and Zhengjun Zhang. I have been very fortunate to be a student of Professor Frees. He has provided endless support and guidance to me, and always believed in me. I appreciate his patience and encouragement when I was struggling. I am blessed to have such a wonderful role model for my career. I am also grateful to Professor Zhang for his insight, support, and advice for my research and career.

I would like to thank Professor Peng Shi, who I collaborated with in my last two years. I found the experience working with him very enjoyable, and he set an outstanding example of how to be a good researcher. I thank Professors Jee-Seon Kim and Chunming Zhang for being on my committee and giving me support and invaluable ideas to complete this dissertation.

I am indebted to all my friends during my time as a graduate student, who have made my life in Madison pleasant and memorable. I also appreciate all the help from the staff in the department.

Finally, I want to thank my family for their support and love. My parents have always encouraged me to pursue my interests. I also want to thank my boyfriend Jared Huling. Thank you for being through all the ups and downs in graduate school with me.

Chapter 1

Introduction

Multivariate discrete outcomes are common in a wide scope of areas, including insurance, psychometrics, and epidemiology. For instance, in property insurance, it is common that a policy contains multiple coverage types, e.g. building and contents coverage and motor vehicle coverage. Hence, insurers can observe claim frequencies, indicating whether or not a claim has occurred or the number of claims, of multiple types from a policyholder. When the interplay between outcomes has significant consequences, modeling dependencies among interrelated variables is of great importance. In the foregoing example, quantifying dependencies among risks is critical for understanding the uncertainty of the portfolio, and thus is important for an insurer's solvency and profitability.

1.1 Models for Multivariate Discrete Outcomes

There are many good approaches available for modeling multivariate discrete outcomes. Generalized linear mixed models (McCulloch and Neuhaus 2001) have been extensively applied to handle correlated discrete observations, though the models do not keep the marginal distributions after integrating out random effects.

There are extensive literatures on multivariate binary outcomes. For d-dimensional binary data, there are 2^d possible combinations of outcomes. Hence, multinomial logistic

regression models can be employed to fit the combinations. Nevertheless, the interpretation of multinomial logistic regression models is not quite straightforward. Frees et al. (2013) used dependence ratio and odds ratio methods, for which the likelihood is written as a function of marginal parameters and odds ratio or dependence ratio. None of these models appears to be uniformly preferable to the others.

The existing literature also contains a variety of models for multivariate counts. One commonly used approach of introducing dependencies among counts is through common additive errors, for instance, a multivariate Poisson model with a common covariance parameter (Johnson et al. 1997; Tsionas 2001). As an example, three-dimensional Poisson outcomes (Y_1, Y_2, Y_3) can be expressed by

$$\begin{cases} Y_1 = X_1 + X_0 \\ Y_2 = X_2 + X_0 \\ Y_3 = X_3 + X_0, \end{cases} \tag{1.1}$$

where $X_j, j \in \{0, 1, 2, 3\}$ independently follows a Poisson distribution with a parameter λ_j, which can be related to covariates. Each pair shares a common variable X_0 that induces a common correlation, say λ_0. Extending this idea, a more flexible way is a model with full covariance (Karlis and Meligkotsidou 2005). Specifically, assume $X_i \sim \text{Poisson}(\lambda_i), i \in \{1, 2, 3\}$ and $X_{ij} \sim \text{Poisson}(\lambda_{ij}), i, j \in \{1, 2, 3\}, i < j$, then

$$\begin{cases} Y_1 = X_1 + X_{12} + X_{13} \\ Y_2 = X_2 + X_{12} + X_{23} \\ Y_3 = X_3 + X_{13} + X_{23}. \end{cases} \tag{1.2}$$

Thus, $Y_i, i \in \{1, 2, 3\}$ follows a Poisson distribution marginally with parameter $\lambda_i + \lambda_{ij} + \lambda_{ik}$, $i, j, k \in \{1, 2, 3\}, i \neq j \neq k$, and the covariance between Y_i and Y_j is λ_{ij}. Multivariate negative binomial distributions built through a mixture distribution (Joe 1997) or a common additive factor (Winkelmann 2000) and zero-inflated multivariate Poisson models (Bermúdez

and Karlis 2011) can be applied in the presence of overdispersion. A limitation of the foregoing models is that they allow only positive correlations. There are models that allow negative correlations, such as multivariate Poisson-log-normal models (Aitchison and Ho 1989), the correlated latent effects approach (Chib and Winkelmann 2012), and the conditional probability approach (Berkhout and Plug 2004). However, for some datasets, different marginal models than the commonly used ones or combinations of different marginal models might be necessary (Frees et al. 2016).

1.2 Applications of Copulas on Discrete Outcomes

This dissertation uses a probabilistic structure known as a *copula*, which is a multivariate distribution function with uniform margins that has been used to study dependencies in many areas including, but not limited to, insurance (Frees and Valdez 1998), finance (Li 1999), and survival analysis (Shih and Louis 1995); see Nelsen (2007) for an introduction. Sklar's Theorem (Sklar 1959) provides a theoretical foundation for copulas as useful tools to connect margins and dependence. For any d dimensional variable $(Y_1, \ldots, Y_d)$ with marginal distribution functions $F_1(\cdot), \ldots, F_d(\cdot)$, there exists at least one copula C such that

$$F(\cdot) = C(F_1(\cdot), \ldots, F_d(\cdot)), \tag{1.3}$$

i.e., the joint distribution can be expressed in terms of margins and a copula. Sklar's theorem is unified over continuous, discrete, and mixture cases. For discrete outcomes taking integer values, parametric copulas can be fit through maximum likelihood estimation (MLE) straightforwardly using (1.3)

$$P(Y_1 = y_1, \ldots, Y_d = y_d) = \sum_{l_1=0}^{1} \cdots \sum_{l_d=0}^{1} (-1)^{l_1 + \cdots + l_d} C\left[F_1(y_1 - l_1), \ldots, F_d(y_d - l_d) \right]. \tag{1.4}$$

There are numerous copula families which can accommodate different dependence structures such as negative correlations, asymmetry, and tail dependence (Joe 1993; Nelsen 2003; Yang

et al. 2011).

Although most applications focus on continuous variables, there is an increasing trend in the application of copulas on discrete outcomes. For binary outcomes, the widely used multivariate probit model (Brown 1998; Young et al. 2009) is indeed a special case of copula regression models using probit margins and a Gaussian copula (Song 2007). For more applications, Genest et al. (2013) studied multivariate logistic models using copulas. They fit multivariate binary variables with meta-elliptical copulas which can be viewed as extensions of Gaussian or t copulas and derived a test of pairwise independence for binary outcomes. Nikoloulopoulos and Karlis (2008) used mixture of max-id copulas which have both global parameters and pairwise parameters to fit multivariate binary data. They also proposed a bootstrap model averaging method for calculating Kendall's τ and indicated the covariates do influence copula parameters.

There are also expanding applications in count data. Nikoloulopoulos and Karlis (2009) applied partially symmetric copulas and mixture of max-id copulas on transaction market basket data to model the purchase frequencies of different product categories. Partially symmetric copulas are extensions of Archimedean copulas with $d - 1$ dependence parameters and more flexible than Archimedean copulas. Kolev and Paiva (2009) applied a nested Archimedean copula to model insurance usages of couples, and Shi and Valdez (2014) modeled claim frequencies from different types of coverage using negative binomial margins and copulas.

1.3 Issues of Copulas on Discrete Outcomes

Yet the application of copulas on discrete data is still in its infancy; one of the biggest barriers is the identifiability of copulas. Sklar (Sklar 1959) showed the uniqueness of copulas is only guaranteed at the range of marginal distribution functions, and thus the copula representation may not be unique for discrete outcomes whose ranges of marginal distribution functions are only a countable number of points.

Genest and Nešlehová (2007) provided a thorough investigation on the non-identifiability issue. Here we use the bivariate case for illustration. Consider a bivariate discrete variable (Y_1, Y_2) with joint distribution function F and marginal distributions F_1 and F_2. The set of copulas compatible with (Y_1, Y_2) denoted as $\mathcal{C}_F$, in the sense that (1.3) holds for any $C \in \mathcal{C}_F$, is generally quite large. Due to lack of uniqueness, modeling and interpreting dependence for discrete outcomes using copulas is subject to caution. For example, independent variables may relate to copulas other than the independence copula. To quantify how large $\mathcal{C}_F$ could be, Genest and Nešlehová (2007) showed for any two copulas in the set, their pointwise distance can be bounded by the maximum marginal point mass (Proposition 6), which indicates non-identifiability is more of a problem when Y_1 or Y_2 takes on a small number of values, e.g. binary variables.

Another issue with copulas for discrete outcomes is that commonly used summary statistics, such as concordance measures, can give misleading results. As an example, Kendall's τ, one of the most common concordance measures, is problematic for discrete outcomes due to the fact that the definition of Kendall's τ does not take the probability of ties into account. As a result, Kendall's τ depends on marginal distribution functions for discrete outcomes. Denuit and Lambert (2005) derived the bounds of Kendall's τ when changing marginal parameters. It was noted that perfect monotone dependence does not necessarily imply concordance measures to be 1. In addition, Kendall's τ of the data based on the probabilistic definitions does not always coincide with the ones of the related copulas based on the analytical definition. As a result, it induces problem in inferences; moment based estimation by inverting sample Kendall's τ lead to a biased estimator of the copula parameter. In general, MLE is more reliable for discrete outcomes.

The existing literature contains works solving the issues of copulas on discrete outcomes from different aspects. Since ties can cause the rank-based estimation methods of copula parameters to fail, Li et al. (2016) studied the estimation of copula parameters in presence of ties in data. Naive approaches are to use average ranks or to break the ties at random. Instead, they used the concept of interval censored data from survival analysis to calculate

the pseudo likelihood. They used bootstrap to construct confidence intervals. Note that parametric bootstrap methods do not work for data with ties as they do not generate ties from fitted parametric copulas, so Li et al. (2016) adjusted it by inverting the empirical distribution of the pseudo observations. They also pointed out that goodness-of-fit tests with standard parametric bootstrap are vulnerable to ties in keeping their sizes.

For an alternative approach, Denuit and Lambert (2005) studied continuous extensions (jittering) defined as

$$Y_j^* = Y_j + (V_j - 1),$$

where V_1 and V_2 independently follow uniform distributions between $(0, 1)$. They derived the form of the unique copula of (Y_1^*, Y_2^*) and its Kendall's τ using the distribution function of (Y_1, Y_2). The generated (Y_1^*, Y_2^*) has some good properties. For example, the copula for (Y_1^*, Y_2^*) belongs to $\mathcal{C}_F$, and Kendall's τ is invariant through jittering, i.e., $\tau(Y_1, Y_2) = \tau(Y_1^*, Y_2^*)$. The concordance order is also preserved through jittering in the sense that for another pair of discrete outcomes (X_1, X_2) with distribution function L,

$$L \prec_c F \Leftrightarrow (X_1^*, X_2^*) \prec_c (Y_1^*, Y_2^*),$$

where $L \prec_c F$ is defined by that $L(x_1, x_2) \leq F(x_1, x_2)$ for all x_1, x_2.

Computational efficiency is another concern for applying copulas on discrete outcomes. Since the likelihood (1.4) involves the computation of rectangle probabilities, the computation could be heavy with elliptical copulas and high dimensions. Numerical methods to compute the multivariate integration for elliptical copulas have been discussed in the literature; see a review in Nikoloulopoulos (2013). Instead of computing high-dimensional integration directly, composite likelihood estimation (Joe 2014) which is based on a pseudo likelihood as weighted summations of marginal densities of lower dimensions leads to efficient computation. Alternatively, Panagiotelis et al. (2012) adapted vine copulas to discrete outcomes using bivariate copulas as building blocks to provide flexible models. The computation is of exponential order for the full copula models, while it is polynomial order using

vine copulas. Hence it improves the computational efficiency significantly. They derived the algorithm for computing the likelihood of vine copula.

1.4 Focus of The Thesis

Copulas are commonly applied in the regression settings in which outcomes are related to a set of covariates (Song et al. 2009; Kolev and Paiva 2009). Copula regression can preserve the solid body of work established for marginal models (e.g., McCullagh and Nelder 1989; Agresti and Kateri 2011) and accommodate dependence structures flexibly. Each marginal distribution can be specified to be conditioned on its covariates. It is customary to assume a dependence structure that is constant over observations. We use this simplifying assumption in this paper for the purposes of easy interpretation; see alternatives in Patton (2006), Nikoloulopoulos and Karlis (2010), and Acar et al. (2011).

As emphasized in the prior section, the non-identifiability issue of copulas on discrete outcomes has concerned analysts, calling into question model interpretations and predictions. We study the issue of identifiability in a regression context. In the i.i.d. case, it is impossible to identify a continuous copula function on the discrete support. The difficulty vanishes if we assume that there exists at least one continuous component of regressors with a nonzero coefficient. As the marginal distributions vary with covariates, inclusion of continuous regressors provides a region of support for copula identifiability. In this dissertation, we provide sufficient conditions under which the copula regression model is identifiable for discrete outcomes.

Given identifiability, how to correctly specify a copula model has remained a question. When each Y_j, $j = 1, \ldots, d$ is continuous, the probability integral transform $F_j(Y_j)$ is uniformly distributed, and the unique underlying copula is actually the joint distribution of $(F_1(Y_1), \ldots, F_d(Y_d))$. Hence copula identification can be conducted using the probability integral transforms, or the pseudo observations, by checking properties such as tail dependence and asymmetry in scatter plots (Joe 2014) or through formal tests (Li and Genton

2013). Section 1.4 of Joe (2014) gives an example of copula specification for continuous outcomes. They computed empirical normal scores which is of the form $\Phi^{-1}[F_j(Y_j)]$ and compared the scatter plot with the elliptical contours of the bivariate normal density from different copula families to identify the features of the underlying copula. For example, sharper corners indicate tail dependence, and the difference between lower and upper semi-correlation reflects asymmetry. The reason they used normal scores instead of uniform margins is that it is better to assess tail dependence and asymmetry as the copula density asymptotes to ∞ for at least one corner for many commonly used bivariate copula families.

Nonetheless, for a discrete outcome such as Y_1, the distribution of $F_1(Y_1)$ is generally not uniform, and the related copulas do not coincide with the joint distribution function of $(F_1(Y_1), \ldots, F_d(Y_d))$. Thus, the approaches of copula specification for continuous outcomes cannot be applied directly to discrete outcomes.

Meanwhile, many informative diagnostic tools in continuous cases are also based on probability integral transforms, for instance their empirical distributions (Deheuvels 1979) and transformations (Rosenblatt 1952; Genest et al. 2009)). As a result of "ill" probability integral transforms, there are few approaches available for diagnosis in discrete cases. The classical way of comparing empirical and observed counts is infeasible when there are many large observations and hard to present when the dimension is greater than 2. In practice, overall goodness-of-fit statistics, such as AIC, BIC, and likelihood, are used to choose the best model among candidates (Nikoloulopoulos and Karlis 2010; Shi and Valdez 2014). Vuong's test (Vuong 1989) can be applied to further compare if the models are statistically significantly different. However, these methods are not diagnostic for adequacy of fit and do not suggest improvements. In addition, AIC and BIC do not necessarily indicate goodness of fit at tail.

To identify the "hidden" dependence structure under discreteness, in this paper, we develop a nonparametric copula estimator. Most existing nonparametric copula estimators (Deheuvels 1979; Chen and Huang 2007; Omelka et al. 2009) assume continuity of the marginal distributions. Instead of applying the empirical estimator, we construct the esti-

mator based on a local average approach. For practitioners who prefer to use parametric copulas, the proposed nonparametric estimator can also serve as a diagnostic and specification tool for selecting a parametric form of copulas. Adequacy of fit can be checked by comparing the fitted parametric copula with the nonparametric estimator.

The rest of the thesis is organized as follows. In Chapter, 2 we propose our "perturbed" probability integral transformation under discreteness. This chapter sets the tone and notation for the rest of the dissertation. In Chapter 3, we present the proposed nonparametric estimator and its asymptotic properties. Chapter 4 contains our simulation studies, and in Chapter 5 we analyze the data from the Wisconsin Local Government Property Insurance Fund.

As ongoing and future work, in Chapter 6, we extend the framework of copula identification to mixed type of data. For example, insurance claim outcomes usually follow a mixed distribution of a point mass at zero with a positive distribution. Due to the discrete component, copula identification remains a problem. We propose a nonparametric copula estimator for mixed data. In Chapter 7, we discuss the usage of the proposed nonparametric estimator for goodness-of-fit test of marginal models as well as copulas. Discussion and conclusions are also included in Chapter 7.

Chapter 2

Univariate Perturbed Probability Integral Transform

Under continuity, copula regression models can be fit through the following stages:

1. Fit marginal models.

2. Transform margins into uniform pseudo observations through probability integral transformations.

3. Specify and estimate copula models using the pseudo observations.

That is, after margins have been fit, a key step is the probability integral transformation, which serves two important functions. On the one hand, it provides a goodness-of-fit test for marginal models. If the pseudo observations do not show uniform trend, it indicates marginal models should be improved. On the other hand, it lays the foundation for copula specification.

Unfortunately, the transformation for continuous outcomes which will be discussed in Section 2.1 does not apply readily to discrete outcomes. In this chapter, we construct a perturbed uniform transformation under our sampling scheme to approximate the process, which lay the foundation for copula identification for Chapter 3. We can also use the

perturbed transformation to build a goodness-of-fit test for marginal models, which will be further discussed in Section 7.2.

2.1 Probability Integral Transform under Continuity

Without loss of generality, take the jth variable of interest Y_j as an example. We suppress subscript j in the current chapter for simplicity. Conditioning on $X = x$, Y follows a distribution function $F(y|x)$, where F depends on parameters β. Note β might contain location, scale, and shape parameters. If Y is continuous, plugging (X, Y) in F, the variable $F(Y|X)$ is known as the probability integral transform. For a fixed value $s \in (0, 1)$,

$$P(F(Y|X) \leq s) = s, \tag{2.1}$$

i.e., $F(Y|X)$ is uniformly distributed.

In practice, let $(X_i, Y_i), i = 1, \ldots, n$ be an i.i.d sample of (X, Y). With a fitted marginal model $\hat{F}$, by plugging the data into the formula of the probability integral transform, we can obtain a sequence of Cox-Snell residuals (Cox and Snell 1968) $\hat{F}(Y_i|X_i), i = 1, \ldots, n$. For a fixed point $s \in (0, 1)$, the empirical distribution of Cox-Snell residuals taking value at s

$$\hat{U}(s; \hat{\beta}) = \frac{1}{n} \sum_{i=1}^{n} 1(\hat{F}(Y_i|X_i) \leq s) \tag{2.2}$$

should be close to s under correct model specification. Thus, it can be used as a diagnostic tool for univariate model fitting; see applications in Shi and Frees (2011).

2.2 Perturbed Empirical Residual Distribution Function

However, when Y is a discrete variable, the Cox-Snell residuals are not uniformly distributed. Figure 2.1 portrays the Cox-Snell residuals of simulated examples. On the left panel, the data were generated with a gamma regression model and the Cox-Snell resid-

uals were calculated using the underlying model. As anticipated, the Cox-Snell residuals appear to be uniform. In contrast, when the data were generated from a logistic regression model, the Cox-Snell residuals are far apart from uniformity even with the knowledge of the underlying model, as shown in the right panel of Figure 2.1.
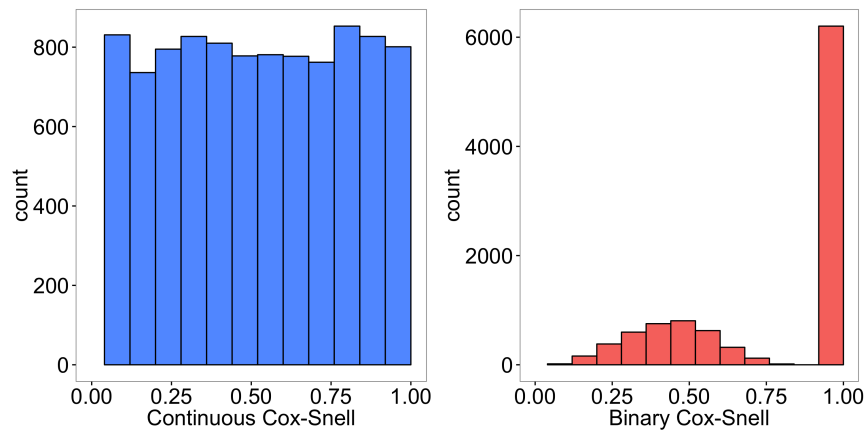
Figure 2.1: Histogram of Cox-Snell residuals. Left panel: gamma regression. Right panel: logistic regression.

Probability integral transformations invalidate under discreteness due to the fact that (2.1) is not true for some values of s. To see where (2.1) does hold, we define the conditional range of the marginal distribution function given X as a grid $\Lambda(X) = \{F(k|X) : k = 0, 1, \ldots\}$. Note that the range of Y can be finite, e.g., binary distribution. Equation (2.1) is true if $s \in \Lambda(X)$.

To construct an alternative to $\hat{U}(\cdot; \beta)$ as in (2.2) under discreteness, ideally, if we could find a subset of observations for which $s \in \Lambda(X)$ holds, those observations can be plugged in (2.2). We require X contains continuous components to achieve copula identifiability, which will be discussed in Section 3.1. When X varies in regression, there might be a subset of observations for which $s \in \Lambda(X)$ holds approximately.

To formalize this idea, we condition on X and denote $F^{(-1)}(\cdot|X)$ as the general inverse function of $F(\cdot|X)$ such that $F^{(-1)}(s|X) = \min\{y : F(y|X) \geq s\}$ for $s \in (0, 1)$. Denote

$$H^+(s; X) = F(F^{(-1)}(s|X)|X).$$

It can be seen that $H^+(s; X) = \min\{\eta \in \Lambda(X) : \eta \geq s\}$, i.e., $H^+(s; X)$ is the smallest point on the grid $\Lambda(X)$ that is larger than or equal to s. In the same way, we can define the largest point on the grid $\Lambda(X)$ that is smaller than or equal to s as

$$H^-(s; X) = \max\{\eta \in \Lambda(X) : \eta \leq s\}.$$

To combine these two cases, define the grid point closest to s, or "perturbed" probability integral transform,

$$H(s; X) = \begin{cases} H^+(s; X) & H^+(s; X) + H^-(s; X) \leq 2s, \\ H^-(s; X) & \text{Otherwise.} \end{cases} \tag{2.3}$$

The distance between s and $\Lambda(X)$ can be quantified by the difference between s and $H(s; X)$. For an observation that s is "close to" being on its grid in the sense that $H(s; X) \approx s$, we can build an approximation to (2.1)

$$P\left(F(Y|X) \leq H(s; X)\right) = H(s; X) \approx s,$$

where the first equation holds due to the fact that $H(s; X) \in \Lambda(X)$ from its definition.

Now consider a sample $(Y_i, X_i), i = 1, \ldots, n$ and a fixed value of s. As s is close to the grid of some observations while not for others, we use a kernel function $K(\cdot)$ to assign weights to observations depending on the normalized distance between s and $H(s; X_i)$ using the form $K\left[(H(s; X_i) - s)/\epsilon_n\right]$ with bandwidth ϵ_n.

The above observation motivates the definition of the perturbed empirical residual distribution function as an alternative to (2.2)

$$\hat{U}(s; \beta) = \sum_{i=1}^n W_{ni}(s; X_i, \beta) 1\left[F(Y_i|X_i) \leq H(s; X_i)\right], \tag{2.4}$$

where

$$W_{ni}(s; X_i, \beta) = \frac{K\left[(H(s; X_i) - s)/\epsilon_n\right]}{\sum_{i=1}^n K\left[(H(s; X_i) - s)/\epsilon_n\right]},$$

and K is a bounded and symmetric kernel. Intuitively, we put large weights on the observations for which s is closely on their grids, while putting small weights otherwise. The term $\hat{U}(s; \beta)$ should be close to s for $s \in (0, 1)$ under true model, hence it can be used as a diagnostic tool.

In practice, β is unknown; let $\hat{\beta}$ be the corresponding estimator. By plugging $\hat{\beta}$ in (2.4), we may obtain $\hat{U}(\cdot; \hat{\beta})$.

As an example, when Y is a binary outcome, its marginal distribution grid only contains two points. Denote the probability of 0 given X as $F(0|X)$, and hence $\Lambda(X) = \{F(0|X), 1\}$. As ϵ_n is small enough, only $F(0|X)$ can possibly be in the small neighborhood of s. Therefore, (2.4) becomes

$$\frac{\sum_{i=1}^n 1\left[|F(0|X_i) - s| \leq \epsilon_n\right] 1(Y_i = 0)}{\sum_{i=1}^n 1\left[|F(0|X_i) - s| \leq \epsilon_n\right]}.$$

This statistic can be recognized as a Nadaraya-Watson estimator and so its asymptotic properties, such as consistency and asymptotic normality, are well established.

In contrast, when Y has an infinite range, for instance Poisson variables, $\hat{U}(s; \beta)$ is a nonstandard estimator in the following aspects. First, for a fixed point s, $H(s; X)$ is a noncontinuous variable. To illustrate, we assume that Y follows the commonly used Poisson generalized linear model (GLM) with the log link, i.e., $Y|X \sim \text{Poisson}(\exp(X'\beta))$. Figure 2.2 shows $H(s; X)$ as a function of $\mu = X'\beta$ (red solid curves). In this example, for a fixed k, $F(k|X)$ is a monotone decreasing function of μ (dashed lines). From Figure 2.2, the curve of $H(s; X)$ is comprised of continuous pieces from the curves of $F(k|X), k = 0, \ldots$. To formalize, denote M^k as the jump point of $H(s; X)$ on the curve of $F(k|X)$, as in Figure 2.2, and $M^0 = -\infty$. For example, when $\mu < M^1$, $F(0|X)$ is closer to s than $F(1|X)$ and $H(s; X) = F(0|X)$ from the definition of $H(s; X)$ that is the closest grid point to s. When $\mu = M^1$, $F(0|X)$ and $F(1|X)$ are equidistant from s. While $\mu > M^1$, $F(1|X)$ is closest to

s and thus $H(s; X) = F(1|X)$. To generalize, it can be seen that

$$H(s; X) = F(k|X) \text{ when } M^k \leq \mu < M^{k+1}. \tag{2.5}$$

Hence, the random variable $H(s; X)$ is a continuous function of μ almost everywhere except at a countable number of points where s is in the middle of two grid points. We will further analyze the issue of discontinuity in Section 3.3. Because of these discontinuities, the proof for asymptotic properties of the estimator is not trivial.
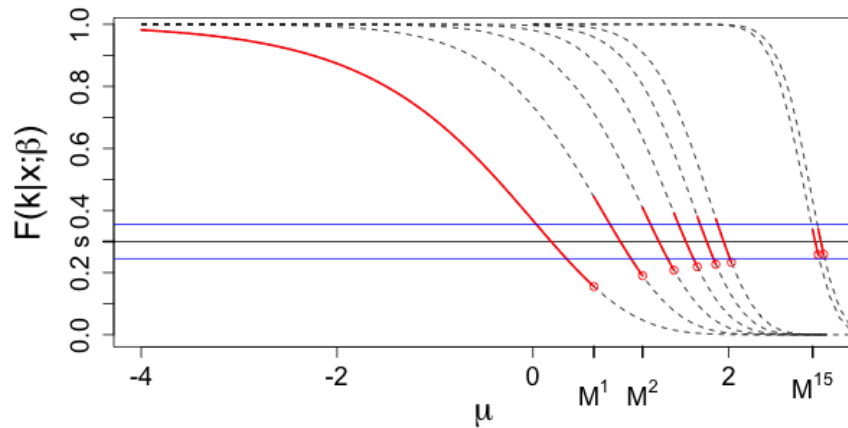
Figure 2.2: $H(s; X)$ (solid red curve) as a function of $\mu = X'\beta$ for Poisson GLM with the log link. Dashed black curves: $F(k|X)$, from left to right $k = 0, 1, 2, 3, 4, 5, 15, 16$. The curve of $H(s; X)$ is comprised of pieces from the curves of $F(k|X), k = 0, \ldots$. Blue horizontal lines: $s + \epsilon, s - \epsilon$.

Second, $H(s; X)$ is a function of s. That is, when we check the estimator at different points, we plug different variables into the kernel function, which also differs from the setting of traditional nonparametric regression models. The dynamic scheme increases efficiency especially when the data are less discrete. This point will be further addressed in the multivariate context in Chapter 3.

Chapter 3

Nonparametric Copula Estimator

In what follows, we focus on the bivariate case first for simplicity. The results can be naturally extended to higher dimensions. Let $\mathbf{Y} = (Y_1, Y_2)'$ be discrete response variables taking integer values with corresponding covariates $\mathbf{X} = (X_1', X_2')'$. Each marginal distribution is specified to be conditioned on its covariates, i.e., for $j = 1, 2$, conditioning on $X_j = x$, Y_j follows a distribution function $F_j(y|x)$, where F_j depends on parameters β_j.

We assume that C does not change with covariates for the purpose of identifiability and easy interpretation, which will be further explained in Section 3.1. Hence, from Sklar's Theorem, there exists a copula C such that

$$F(\cdot|\mathbf{x}) = C\left(F_1(\cdot|x_1), F_2(\cdot|x_2)\right). \tag{3.1}$$

The goal is to identify C.

3.1 Identifiability

The first question concerns identifiability, i.e., whether C can be uniquely determined by the population distribution of $(\mathbf{X}, \mathbf{Y})$. This issue has been addressed in Genest and Nešlehová (2007) in the setting without regressors $\mathbf{X}$. It was shown by Sklar that the copula is only unique over $\mathrm{Ran}(F_1) \times \mathrm{Ran}(F_2)$, where $\mathrm{Ran}(F_j)$ denotes the range of F_j. The copula

functions that equate $C\left(F_1(k_1), F_2(k_2)\right)$ to $F(k_1, k_2)$, for $(k_1, k_2)'$ taking the possible values of $\mathbf{Y}$, are compatible with the data, which only constrains the copula on a discrete number of points. There are infinitely many such copulas that are observationally identical and would be indistinguishable from one another even with the knowledge of the population distribution of $\mathbf{Y}$. As a most extreme example, for bivariate binary outcomes, we are only able to identify the copula at the point $(F_1(0), F_2(0))$.

The non-identifiability issue of copulas on discrete outcomes has concerned analysts. First, the qualified copulas have different properties such as Kendall's τ and tail dependencies, which results in difficulties of interpretation. Second, one may want to make predictions outside the range of observations; identifiability is essential for extrapolation.

In the regression setting, in contrast, for a fixed integer k_j, $F_j(k_j|x_j)$ is a function of x_j. For example, for logistic regression models, $F_j(0|x_j) = 1/\left[1 + \exp(x_j'\beta_j)\right]$. Hence, inclusion of continuous covariates widens the range of $F_j(k_j|x_j)$ from a discrete number of points to an interval. Together with the assumption that the copula does not change with covariates, the copula function can be uniquely determined by the population at the region composed of possible values of $(F_1(k_1|x_1), F_2(k_2|x_2))$.

3.2 Perturbed Empirical Copula Estimator

Given identifiability of the copula over a region, now we focus on how to correctly specify the model. If Y_1 and Y_2 are continuously distributed, jointly, for a fixed point $(s, t) \in (0, 1)^2$, from (3.1),

$$P\left(F_1(Y_1|x_1) \le s, F_2(Y_2|x_2) \le t|\mathbf{x}\right) = F\left(F_1^{(-1)}(s|x_1), F_2^{(-1)}(t|x_2)|\mathbf{x}\right) = C(s, t),$$

and hence it can be seen that

$$P\left(F_1(Y_1|X_1) \le s, F_2(Y_2|X_2) \le t\right) = \mathrm{E}\left[P\left(F_1(Y_1|X_1) \le s, F_2(Y_2|X_2) \le t|\mathbf{X}\right)\right] = C(s, t).$$

$$(3.2)$$

That is, the copula related to $\mathbf{Y}$ is the joint distribution function of $(F_1(Y_1|X_1), F_2(Y_2|X_2))$. Equation (3.2) is essential for copula identification and estimation under continuity.

In practice, let $(\mathbf{X}_i, \mathbf{Y}_i)$, $i = 1, \ldots, n$ be an i.i.d sample of $(\mathbf{X}, \mathbf{Y})$. The empirical distribution of the bivariate Cox-Snell residuals $\left(\hat{F}_1(Y_{i1}|X_{i1}), \hat{F}_2(Y_{i2}|X_{i2})\right)$

$$\frac{1}{n} \sum_{i=1}^{n} 1\left(\hat{F}_1(Y_{i1}|X_{i1}) \leq s, \hat{F}_2(Y_{i2}|X_{i2}) \leq t\right) \tag{3.3}$$

(Deheuvels 1979) or its kernel estimator (e.g., Scaillet and Fermanian 2002; Chen and Huang 2007) provides nonparametric estimation of C.

However, when Y_1 and Y_2 are discrete, similar to the univariate case, (3.2) does not always hold, and thus the nonparametric copula estimators for continuous outcomes do not readily apply. In Chapter 2, we introduced the perturbed empirical residual distribution function. Since a copula is a multivariate distribution with uniform margins, it is natural to extend the approach in Chapter 2 to copulas. To see where (3.2) does hold, we define the conditional range of the distribution function given $\mathbf{X}$ as a two-dimensional grid $\Lambda(\mathbf{X}) = \{(F_1(k_1|X_1), F_2(k_2|X_2)), k_1 = 0, 1, \ldots, k_2 = 0, 1, \ldots\}$. It can be seen that $\Lambda(\mathbf{X}) = \Lambda_1(X_1) \times \Lambda_2(X_2)$. Equation (3.2) is true if when (s, t) is on the grid $\Lambda(\mathbf{X})$.

To formalize this idea, we denote $H(s, t; \mathbf{X}) = (H_1(s; X_1), H_2(t; X_2))$, where the marginal perturbed probability integral transforms $H_1(\cdot; X_1)$, $H_2(\cdot; X_2)$ are defined in (2.3). It can be seen that $H(s, t; \mathbf{X}) \in \Lambda(\mathbf{X})$ and is the closet grid point to (s, t). For an observation that (s, t) is "close to" being on its grid in the sense that $H_1(s; X_1) \approx s, H_2(t; X_2) \approx t$, we can build an approximation to (3.3) due to the fact

$$P\left(F_1(Y_1|X_1) \leq H_2(s; X_1), F_2(Y_2|X_2) \leq H_2(t; X_2)\right) = C\left(H_2(s; X_1), H_2(t; X_2)\right) \approx C(s, t),$$

where the first equation holds since $H(s, t; \mathbf{X}) \in \Lambda(\mathbf{X})$ from its definition.

Motivated by the univariate case, we can estimate the copula using the subset for which (s, t) is on the grid $\Lambda(\mathbf{X}_i)$ approximately, or equivalently, $|H(s, t; \mathbf{X}_i) - (s, t)| \leq \epsilon_n$. If the

copula is smooth enough, it is approximately constant over a small enough neighborhood. Recall that $\beta = (\beta_1, \beta_2)$ is the underlying marginal parameter. For simplicity, denote

$$Y_i(\beta) = 1\left[F_1(Y_{i1}|X_{i1}) \leq H_1(s; X_{i1}), F_{i2}(Y_{i2}|X_{i2}) \leq H_2(t; X_{i2})\right]. \tag{3.4}$$

Hence, the copula estimator is

$$\hat{C}(s, t; \beta) = \sum_{i=1}^{n} W_{ni}(s, t; \mathbf{X}_i, \beta) Y_i(\beta), \tag{3.5}$$

where

$$W_{ni}(s, t; \mathbf{X}_i, \beta) = \frac{K((H_1(s; X_{i1}) - s)/\epsilon_n, (H_2(t; X_{i2}) - t)/\epsilon_n)}{\sum_{i=1}^{n} K((H_1(s; X_{i1}) - s)/\epsilon_n, (H_2(t; X_{i2}) - t)/\epsilon_n)},$$

and K is a bounded and symmetric kernel.

In practice, β is unknown; let $\hat{\beta}$ be the corresponding estimator. By plugging $\hat{\beta}$ in (3.5), we may obtain the copula estimator $\hat{C}(s, t; \hat{\beta})$. It will be shown in the following section that the uncertainty in the coefficients is negligible under mild regularity conditions.

3.3   Asymptotic Behavior

In this section, we study the asymptotic properties of the copula estimator $\hat{C}(s, t; \hat{\beta})$ defined in (3.5). We first analyze $\hat{C}(s, t; \beta)$, then plug in the estimator of β.

As mentioned in Chapter 2, marginally, for $j = 1, 2$, for a fixed k, $F_j(k|X_j)$ is a monotone decreasing function of a random location parameter $\mu_j = X_j'\beta_j$. Hence, it is random with its own distribution function and, assuming continuity, a density. Let $f_{F_j(k|X_j)}$ denote the density of $F_j(k|X_j)$, and $f_{H_j(s;X_j)}$ is the density of $H_j(s; X_j)$. The value of $f_{H_j(s;X_j)}$ at s, by transformation of random variables, is

$$f_{H_j(s;X_j)}(s) = \sum_{k=0}^{\infty} f_{F_j(k|X_j)}(s).$$

Note the summation may be finite, for example, as with binary outcomes.

We illustrated in Chapter 2 that $H_j(s; X_j)$ is a noncontinuous function of μ_j (Figure 2.2). Here we demonstrate that its density $f_{H_j(s;X_j)}$ is not continuous at s by showing $f_{H_j(s;X_j)}(s + \epsilon)$ has a different form for $\epsilon \neq 0$. Now, extend the notation in (2.5) with subscripts referring to different margins, that is, M_j^k is the jump point of $H_j(s; X_j)$ on the curve of $F_j(k|X_j)$. Given $\mu_j = M_j^k$, denote the corresponding function value of $H_j(s; X_j)$ as ν_j^k.

For a small k such as $k \leq 5$ in Figure 2.2, $|\nu_j^k - s| > \epsilon$, i.e., the jump point of $H_j(s; X_j)$ on the curve of $F_j(k|X_j)$ is outside the ϵ neighborhood of s. Hence, $f_{F_j(k|X_j)}$ contributes to $f_{H_j(s;X_j)}$ at $s + \epsilon$ when applying transformation of random variables. While for large k such as $k = 15$, $|\nu_j^k - s| < \epsilon$, and thus the density of $F_j(15|X_j)$ does not contribute to the density of $f_{H_j(s;X_j)}$ at $s + \epsilon$. Therefore, in this example,

$$\sum_{k=0}^{5} f_{F_j(k|X_j)}(s + \epsilon) \leq f_{H_j(s;X_j)}(s + \epsilon) < \sum_{k=0}^{\infty} f_{F_j(k|X_j)}(s + \epsilon). \tag{3.6}$$

That is, the density $f_{H_j(s;X_j)}$ is not continuous at s due to loss of $f_{F_j(k|X_j)}$ curves contributing to $f_{H_j(s;X_j)}$ at $s + \epsilon$.

Now ϵ takes a small value ϵ_n which goes to 0 as n goes to infinity. The following lemma guarantees the summation on the left of (3.6) can be up to a large number $a_n(s)$ going to ∞, and $f_{H_j(s;X_j)}$ can be approximated by $\sum_{k=0}^{a_n(s)} f_{F_j(k|X_j)}$, which is continuous at s. Therefore, the discontinuity of $f_{H_j(s;X_j)}$ is negligible.

Lemma 3.3.1. *There exists a sequence $a_n(s)$ going to infinity such that for all $k \leq a_n(s)$, $|\nu_j^k - s| > \epsilon_n$, and thus $f_{H_j(s;X_j)}(s + \epsilon_n) \geq \sum_{k=0}^{a_n(s)} f_{F_j(k|X_j)}$.*

The proof of Lemma 3.3.1 can be found in Section 3.5.1. Since there are countable jump points of $H_1(s; X_1)$ and $\epsilon_n \to 0$, Lemma 3.3.1 can be satisfied by choosing right order of $a_n(s)$. Specifically, Lemma 3.3.1 is satisfied by choosing $a_n(s)$ for Poisson GLMs to be $1/\epsilon_n^2$ and $1/\epsilon_n$ for negative binomial distributions; see proofs in Section 3.5.1.

It is worth mentioning that if Y_j has a finite support such as binary outcomes, when ϵ_n is small enough, the jump points can be excluded from the ϵ_n neighbor of s, hence $a_n(s)$

can be chosen as an arbitrarily large number.

Extending our notation, denote the joint density of $(F_1(k_1|X_1), F_2(k_2|X_2))$ as $f_{F_1(k_1|X_1), F_2(k_2|X_2)}$, and $f_{H(s,t;\mathbf{X})}$ as the density of $H(s,t;\mathbf{X})$. It can be seen

$$f_{H(s,t;\mathbf{X})}(s,t) = \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} f_{F_1(k_1|X_1), F_2(k_2|X_2)}(s,t). \tag{3.7}$$

Assume we can interchange the derivatives and the limits, the partial derivatives of $f_{H(s,t;\mathbf{X})}(s,t)$ are $f_{H(s,t;\mathbf{X}),1} = \partial f_{H(s,t;\mathbf{X})}/\partial s$ and $f_{H(s,t;\mathbf{X}),2} = \partial f_{H(s,t;\mathbf{X})}/\partial t$. To show the asymptotic properties of the copula estimator, the following regularity assumption guarantees that $f_{F_1(k_1|X_1), F_2(k_2|X_2)}$ is sufficiently smooth.

Assumption 3.3.1. *For fixed k_1 and k_2, $f_{F_1(k_1|X_1), F_2(k_2|X_2)}$ is twice continuously differentiable. The density $f_{H(s,t;\mathbf{X})}$ and its derivatives are bounded.*

A necessary condition for Assumption 3.3.1 is that there exists a continuous regressor whose coefficient is not 0. Recall $M_1^{a_n(s)}$ denotes the jump point of $H_1(s; X_1)$ on the curve of $F_1(a_n(s)|X_1)$ as in (2.5). The following assumption guarantees the discontinuity of $f_{H(s,t;\mathbf{X})}$ is negligible.

Assumption 3.3.2. *Let $a_n(s)$ and $b_n(t)$ be sequences as in Lemma 3.3.1 for $F_1(k_1|X_1)$ and $F_2(k_2|X_2)$, respectively, then $\epsilon_n^{-2} P\left(\mu_1 > M_1^{a_n(s)}\right) \to 0$ and $\epsilon_n^{-2} P\left(\mu_2 > M_2^{b_n(t)}\right) \to 0$.*

When Y_1 and Y_2 follow Poisson distributions with means $\lambda_1 = \exp(\mu_1)$ and $\lambda_2 = \exp(\mu_2)$, Assumptions 3.3.1 and 3.3.2 are satisfied if $E\lambda_1$, $E\lambda_2$, and $E\sqrt{\lambda_1 \lambda_2}$ are finite, and they hold for negative binomial distributions if $E\lambda_1^2$, $E\lambda_2^2$, and $E\lambda_1 \lambda_2$ are finite. See Propositions 3.4.1 and 3.4.2 for verification. Therefore, if there are highly right-skewed covariates, log transformation is suggested.

The copula is assumed to satisfy smoothness conditions, which guarantees the eligibility of approximating the copula value at a point using its neighborhood.

Assumption 3.3.3. *The copula C for $Y_1|X_1$ and $Y_2|X_2$ does not change with $\mathbf{X}$. The copula is twice continuously differentiable, and the corresponding partial derivatives are*

bounded.

The first part of Assumption 3.3.3 is called "simplifying assumption" in Haff et al. (2010) and Joe (2014). Let V be a subset of $(0,1)^2$ such that for $(s,t) \in V$, $f_{H(s,t;\mathbf{X})}(s,t) > 0$. Denote C_1, C_2, C_{11}, and C_{22} as first and second order partial derivatives of C. Let the bandwidth ϵ_n satisfy that $\epsilon_n \to 0$ and $n\epsilon_n^2 \to \infty$ as $n \to \infty$, and $n\epsilon_n^6 = O(1)$. Assume K is a symmetric and compact supported kernel function. Denote $R_2(K) = \int K(u,v)^2 \mathrm{d}u\mathrm{d}v$, $\kappa_2 = \int u^2 K(u,v)\mathrm{d}u\mathrm{d}v$.

Theorem 3.3.1 (Consistency). *Under Assumptions 3.3.1, 3.3.2, and 3.3.3, for $(s,t) \in V$,*

$$\hat{C}(s,t;\beta) \to_p C(s,t).$$

With further assuming C satisfies Lipschitz condition, we can have normality and corresponding order of $\hat{C}(s,t;\beta)$.

Assumption 3.3.4. *C satisfies Lipschitz condition of order 2, i.e., there exists a constant α_1 such that for any $(a,b),(s,t)$,*

$$|C(s,t) - C(a,b)| \leq \alpha_1 |(s,t) - (a,b)|^2.$$

Theorem 3.3.2 (Normality). *Under Assumptions 3.3.1, 3.3.2, 3.3.3, and 3.3.4, for $(s,t) \in V$,*

$$\sqrt{n\epsilon_n^2} \left(\hat{C}(s,t;\beta) - C(s,t) - \kappa_2 \zeta(s,t)\epsilon_n^2 \right) \to_d N\left(0, \frac{R_2(K)C(s,t)(1-C(s,t))}{f_{H(s,t;\mathbf{X})}(s,t)} \right)$$

where

$$\begin{aligned}
\zeta(s,t) =& \frac{C_{11}(s,t)}{2} + \frac{C_{22}(s,t)}{2} + \frac{C_1(s,t)f_{H(s,t;\mathbf{X}),1}(s,t)}{f_{H(s,t;\mathbf{X})}(s,t)} \\
&+ \frac{C_2(s,t)f_{H(s,t;\mathbf{X}),2}(s,t)}{f_{H(s,t;\mathbf{X})}(s,t)}.
\end{aligned} \tag{3.8}$$

Therefore, the asymptotic mean squared error (AMSE), a commonly used measure of the quality of an estimator, for $\hat{C}(\cdot; \beta)$ at (s, t) is

$$\text{AMSE}\left(\hat{C}(s, t; \beta)\right) = \kappa_2^2 \zeta(s, t)^2 \epsilon_n^4 + \frac{C(s, t)(1 - C(s, t))R_2(K)}{n\epsilon_n^2 f_{H(s,t;\mathbf{X})}(s, t)},$$

which converges to 0.

In principle, the copula can be identified using only the probability of $(0, 0)$ due to the fact $F(0, 0|X_1, X_2) = C(F_1(0|X_1), F_2(0|X_2))$. Assuming that the explanatory variables have sufficient range so that the probabilities of zeros span the interval $(0, 1)$, the corresponding estimator is of the form

$$\hat{C}_0(s, t; \beta) = \frac{\sum_{i=1}^n K\left[(F_1(0|X_{i1}) - s)/\epsilon_n, (F_2(0|X_{i2}) - t)/\epsilon_n\right] 1(Y_{i1} = 0, Y_{i2} = 0)}{\sum_{i=1}^n K\left[(F_1(0|X_{i1}) - s)/\epsilon_n, (F_2(0|X_{i2}) - t)/\epsilon_n\right]}, \quad (3.9)$$

which is an application of the Nadaraya-Watson estimator. From its established asymptotic results (Hansen 2009), the variance of $\hat{C}_0(s, t; \beta)$ is of the form

$$\frac{C(s, t)\left[1 - C(s, t)\right]R_2(K)}{n\epsilon_n^2 f_{F_1(0|X_1), F_2(0|X_2)}(s, t)}.$$

Theorem 3.3.2 underscores the benefit of employing our nonstandard estimator. From the form of $f_{H(s,t;\mathbf{X})}(s, t)$ in (3.7), we can see $f_{H(s,t;\mathbf{X})}(s, t) > f_{F_1(0|X_1), F_2(0|X_2)}(s, t)$. Thus, we have smaller variance by applying the proposed estimator. Intuitively, instead of applying a variable $(F_1(0|X_1), F_2(0|X_2))$, we use many variables $(F_1(k_1|X_1), F_2(k_2|X_2))$, $k_1 = 0, \ldots, k_2 = 0, \ldots$ for copula estimation, which increases the efficiency. This point will be illustrated through a simulation study in Chapter 4.

The following assumptions guarantee the asymptotics when we plug in estimates of the marginal models into the copula estimator $\hat{C}(s, t; \hat{\beta})$. When the parameters are set to be θ, we denote $H(s, t; \mathbf{X}, \theta)$ as the corresponding perturbed probability integral transform.

Assumption 3.3.5 (Lipschitz condition). *There exists a constant α_2 such that for all i,*

for bounded θ and β, when $|\theta - \beta|$ is small enough,

$$|H(s,t;\mathbf{X}_i,\theta) - H(s,t;\mathbf{X}_i)| \leq \alpha_2 |\theta - \beta|$$

almost surely.

Note that this assumption is satisfied when Y_1 and Y_2 follow Poisson GLMs with the log link and bounded covariates; see Section 3.4.1.

Assumption 3.3.6. $n^{1/2}(\hat{\beta} - \beta) = O_p(1)$. *That is, for any $\xi > 0$ there exists $\gamma_\xi > 0$ such that for n big enough,*

$$P(\hat{\beta} \notin B(\beta, n^{-1/2}\gamma_\xi)) < \xi$$

where $B(\beta, d)$ is a neighborhood of β with radius d.

Theorem 3.3.3. *With Assumptions 3.3.1, 3.3.2, 3.3.3, 3.3.4, 3.3.5, and 3.3.6, for $(s,t) \in V$*

$$\sqrt{n\epsilon_n^2}\left(\hat{C}(s,t;\hat{\beta}) - C(s,t) - \kappa_2\zeta(s,t)\epsilon_n^2\right) \to_d N\left(0, \frac{C(s,t)(1-C(s,t))R_2(K)}{f_{H(s,t;\mathbf{X})}(s,t)}\right).$$

That is, the $AMSE$ of the copula estimator $\hat{C}(s,t;\hat{\beta})$ is same as when the margins are known.

Here are a few comments on the asymptotic results. First, the estimator behaves well with large marginal means under which $f_{H(s,t;\mathbf{X})}(s,t)$ is large. For illustration in one dimension, we use Figure 2.2 as an example, and without loss of generality, we focus on the fixed point s in the figure. When μ takes a small value, for instance -3, $H_j(s;X_j) = F_j(0|X_j)$ which is around 1 and not in a small neighborhood of s. While μ is larger, it is more likely that $H_j(s;X_j)$ is in a small neighborhood of s. That is, the density $f_{H_j(s;X_j)}(s)$ is large when μ is mostly distributed at large values. Intuitively, as μ gets large, the grid gets dense and the variable is more similar to a continuous random variable. Extending to two dimensions, when both margins have large means, the estimator performs well. We will verify this point in Chapter 4 with simulated data.

Second, when s is small, it requires large μ value for $F_j(k|X_j)$ to be in a small neighborhood of s. Since we constrain the tail probability of μ by Assumption 3.3.2, $f_{H_j(s;X_j)}(s)$ is small in this case by (3.7). In other words, we have more effective observations when we estimate the copula at the right upper corner than the lower left corner.

Third, here we only provide the results for the bivariate case. The estimator can be extended to higher dimensions naturally with smaller order of convergence $\sqrt{n\epsilon_n^d}$, where d is the number of dimensions. An alternative is to build up the multivariate models through bivariate blocks, known as vine copula structure (Bedford and Cooke 2002; Aas et al. 2009; Panagiotelis et al. 2012).

3.4 Proofs

3.4.1 Verification of Assumptions for Poisson and Negative Binomial Distributions

Proposition 3.4.1. *Assumptions 3.3.1 and 3.3.2 hold for Poisson GLMs with the log link if* $\mathrm{E}\lambda_1$, $\mathrm{E}\lambda_2$, *and* $\mathrm{E}\sqrt{\lambda_1\lambda_2} < \infty$, *where* $\lambda_j = \exp(X_j'\beta_j)$.

Proof. Let Y_1 and Y_2 follow Poisson GLMs. For a fixed k, the distribution function at k given λ_j is

$$F_j(k|\lambda_1) = g(\lambda_j; k) = e^{-\lambda_j} \sum_{i=0}^{k} \frac{1}{i!}\lambda_j^i,$$

which is a smooth monotonically decreasing function of λ_j, as illustrated in Figure 2.2. Hence the derivative of $g(\lambda_j; k)$ with respect to λ_j exists and is of the form

$$g'(\lambda_j; k) = \frac{1}{\partial \lambda_j}\partial g(\lambda_j; k) = -\frac{1}{k!}\lambda_j^k e^{-\lambda_j},$$

which is negative and consistent with the decreasing trend. Due to monotonicity, the inverse function of $g(\lambda_j; k)$ with respect to λ_j exists and is denoted as $g^{-1}(s; k) = \{\lambda : F(k|\lambda) = s\}$. Let f_λ be the density of λ. By the change of variables formula, the density of

$(F_1(k_1|\lambda_1), F_2(k_2|\lambda_2))$ at a point $(s,t) \in (0,1)^2$ can be expressed as

$$f_{F_1(k_1|\lambda_1),F_2(k_2|\lambda_2)}(s,t)$$
$$= f_\lambda \left(g^{-1}(s;k_1), g^{-1}(t;k_2) \right) \frac{1}{g'(g^{-1}(s;k_1);k_1)} \frac{1}{g'(g^{-1}(t;k_2);k_2)}.$$

Thus, using (3.7), the density of $H(s,t;\mathbf{X},\beta)$ at (s,t) is

$$f_{H(s,t;\mathbf{X},\beta)}(s,t) = \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} f_\lambda \left(g^{-1}(s;k_1), g^{-1}(t;k_2) \right) \frac{1}{g'\left(g^{-1}(s;k_1);k_1\right) g'\left(g^{-1}(t;k_2);k_2\right)}.$$

$$(3.10)$$

We now show $f_{H(s,t;\mathbf{X},\beta)}(s,t) < \infty$ under our assumption.

By applying the mean-value forms of the remainder for Taylor series on the exponential function, for any λ_1,

$$e^{\lambda_1} = \sum_{i=0}^{k_1} \frac{1}{i!} \lambda_1^i + \frac{1}{(k_1+1)!} \lambda_1^{k+1} e^{\zeta_1}, \qquad (3.11)$$

where ζ_1 is between 0 and λ_1. Plugging $\lambda_1 = g^{-1}(s;k_1)$ into (3.11) yields

$$1 - s = \frac{1}{(k_1+1)!} g^{-1}(s;k_1)^{k_1+1} e^{\zeta_1} e^{-g^{-1}(s;k_1)}.$$

Taking logarithm on both sides, we get

$$\log\left[(k_1+1)!\right] = (k_1+1)\log\left[g^{-1}(s;k_1)\right] + \zeta_1 - g^{-1}(s;k_1) - \log(1-s).$$

By Stirling's approximation,

$$\log\left[(k_1+1)!\right] \approx (k_1+1)\log(k_1+1) - (k_1+1) + O\left[\log(k_1+1)\right].$$

By comparing the two equations above, we obtain that $g^{-1}(s;k_1)$ is of same order as k_1. Similarly, $g^{-1}(t;k_2)$ is of same rate as k_2.

In addition, by Stirling's approximation, $g'(\lambda_j; k_j)$ given $\lambda_1 = g^{-1}(s; k_1)$ is that

$$-e^{-g^{-1}(s;k_1)}g^{-1}(s;k_1)^{k_1}\frac{1}{k_1!} \sim -\frac{1}{\sqrt{k_1}}.$$

To verify Assumption 3.3.1, by plugging $g^{-1}(s; k_j)$ and $g'(\lambda_j; k_j)$ into (3.10), we have

$$f_{H(s,t;\mathbf{X},\beta)}(s,t) \sim \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \sqrt{k_1 k_2} f_\lambda(k_1, k_2)$$

which converges since $E\sqrt{\lambda_1 \lambda_2} < \infty$.

Now we verify Assumption 3.3.2. We first check the orders of $a_n(s)$ and $b_n(t)$. Note that the probability of $Y_1 = k_1$ given $\lambda_1 = g^{-1}(s; k_1)$ is

$$p_{k_1} = e^{-g^{-1}(s;k_1)}g^{-1}(s;k_1)^{k_1}\frac{1}{k_1!} \sim \frac{1}{\sqrt{k_1}}.$$

In order to find $a_n(s)$ such that $p_{k_1} > \epsilon_n$ for $k_1 \leq a_n(s)$, we can choose $a_n(s)$ and $b_n(t)$ to be of order $1/\epsilon_n^2$. The probability

$$P\left(\lambda_1 > M_1^{a_n(s)}\right) \sim P\left(\lambda_1 > 1/\epsilon_n^2\right),$$

which is $o(\epsilon_n^2)$ since $\mathrm{E}(\lambda_1) < \infty$. Similarly, when $\mathrm{E}(\lambda_2) < \infty$, $1/\epsilon_n^2 P\left(\mu_2 > \mu_2^{b_n(t)}\right) \to 0$. Hence, Assumption 3.3.2 is satisfied. $\qquad\square$

Assumption 3.3.5 is satisfied for Poisson GLM with the log link function since the first order partial derivatives of $F_j(k; \lambda_j)$ with respect to λ_j is bounded by 1. With probability one, when the distance between β and θ is small enough, $H_j(s; X_j, \beta_j)$ and $H_j(s; X_j, \theta)$ are on $F_j(k; \exp(X_j'\beta_j))$ and $F_j(k; \exp(X_j'\theta_j))$ with the same k value.

Proposition 3.4.2. *Assumptions 3.3.1 and 3.3.2 hold for negative binomial GLM with log link if* $\mathrm{E}(\lambda_1^2)$, $\mathrm{E}(\lambda_2^2)$, *and* $\mathrm{E}(\lambda_1 \lambda_2) < \infty$.

Proof. Similar to the proof of Proposition 3.4.1, for a negative binomial distribution with mean parameter λ_j and size parameter α_j. The distribution function at a fixed integer k is

that

$$F(k|\lambda_j, \alpha_j) = g(\lambda_j; k, \alpha_j) = \sum_{i=0}^{k} \binom{i + \alpha_j - 1}{i} \frac{\lambda_j^i \alpha_j^{\alpha_j}}{(\lambda_j + \alpha_j)^{i + \alpha_j}}. \tag{3.12}$$

Denote the derivative of $g(\lambda_j; k, \alpha_j)$ with respect to λ_j as

$$g'(\lambda_j; k, \alpha_j) = \frac{\partial g(\lambda_j; k, \alpha_j)}{\partial \lambda_j} = \frac{\alpha_j}{\lambda_j + \alpha_j} \sum_{i=0}^{k} \left(\left(\frac{i}{\lambda_j} - 1 \right) \binom{i + \alpha_j - 1}{i} \left(\frac{\lambda_j}{\lambda_j + \alpha_j} \right)^i \left(\frac{\alpha_j}{\lambda_j + \alpha_j} \right)^{\alpha_j} \right).$$

Now we check the order of $g^{-1}(s; k_1, \alpha_j)$. It is known that the Maclaurin series for function $g(z) = 1/(1 + z)^{\beta}$ with the mean-value forms of remainder is that

$$\frac{1}{(1 - z)^{\beta + 1}} = \sum_{i=0}^{k} \binom{i + \beta}{i} z^i + \binom{k + \beta + 1}{k + 1} z^{k+1} \frac{1}{(1 - \zeta)^{\beta + k + 2}}, \tag{3.13}$$

where ζ is between 0 and z. Now in (3.13), substituting $z = \frac{\lambda_j}{\lambda_j + \alpha_j}$ and $\beta = \alpha_j - 1$, we have

$$1 = \sum_{i=0}^{k} \binom{i + \alpha_j - 1}{i} \left(\frac{\lambda_j}{\lambda_j + \alpha_j} \right)^i \left(\frac{\alpha_j}{\lambda_j + \alpha_j} \right)^{\alpha_j} +$$
$$\binom{k + \alpha_j}{k + 1} \left(\frac{\lambda_j}{\lambda_j + \alpha_j} \right)^{k+1} \left(\frac{\alpha_j}{\lambda_j + \alpha_j} \right)^{\alpha_j} \frac{1}{(1 - \zeta)^{\alpha_j + k + 1}}.$$

When $\lambda_j = g^{-1}(s; k, \alpha_j)$, we have

$$1 - s = \binom{k + \alpha_j}{k + 1} \left(\frac{\lambda_j}{\lambda_j + \alpha_j} \right)^{k+1} \left(\frac{\alpha_j}{\lambda_j + \alpha_j} \right)^{\alpha_j} \frac{1}{(1 - \zeta)^{\alpha_j + k + 1}}. \tag{3.14}$$

Applying Stirling's approximation on $\binom{k + \alpha_j}{k + 1}$ and taking logarithm, we can have

$$(k + \alpha_j) \log(k + \alpha_j) - (k + 1) \log(k + 1) + \alpha_j - 1 = \log(1 - s) - (k + 1) \log(\lambda_j) - \alpha_j \log(\alpha_j)$$
$$+ (k + \alpha_j + 1) \log(\lambda_j + \alpha_j).$$

$$\tag{3.15}$$

Therefore, comparing the two sides of equation (3.15) yields that $g^{-1}(s; k, \alpha_j)$ is of order k.

When λ_j is of order k and k is large, $\sum_{i=0}^{k} \left(\frac{i}{\lambda_j} \binom{i + \alpha_j - 1}{i} \left(\frac{\lambda_j}{\lambda_j + \alpha_j} \right)^i \left(\frac{\alpha_j}{\lambda_j + \alpha_j} \right)^{\alpha_j} \right)$ is bounded away from 1, hence $\sum_{i=0}^{k} \left(\left(\frac{i}{\lambda_j} - 1 \right) \binom{i + \alpha_j - 1}{i} \left(\frac{\lambda_j}{\lambda_j + \alpha_j} \right)^i \left(\frac{\alpha_j}{\lambda_j + \alpha_j} \right)^{\alpha_j} \right)$ is bounded above zero. Therefore, $g'(\lambda_j; k, \alpha_j)$ is of order $1/k$.

Thus plugging right order of $g'(\lambda_j; k, \alpha_j)$ and $g^{-1}(s; k_1, \alpha_j)$ into (3.10), we can get

$$f_{H(s,t;\mathbf{X},\beta)}(s,t) = \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} f_{F_1(k_1|\lambda_1), F_2(k_2|\lambda_2)}(s,t) \sim \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} f_\lambda(k_1, k_2) k_1 k_2,$$

which converges since $E\lambda_1 \lambda_2 < \infty$.

Now we verify Assumption 3.3.2. We first choose the order of $a_n(s)$. The point mass function when $\lambda_j = k$ is that

$$\binom{k + \alpha_j - 1}{k} \frac{\lambda_j^k \alpha_j^{\alpha_j}}{(\lambda_j + \alpha_j)^{k + \alpha_j}} \sim \frac{\sqrt{k + \alpha_j - 1}}{\sqrt{k}} \left(1 + \frac{\alpha_j - 1}{k} \right)^k e^{-(\alpha_j - 1)} \left(\frac{k + \alpha_j - 1}{k + \alpha_j} \right)^{\alpha_j - 1}.$$

$$\left(1 + \frac{\alpha_j}{k} \right)^k \frac{1}{k + \alpha_j}$$

$$\sim \frac{1}{k + \alpha_j}.$$

Hence we can choose $a_n(s)$ and $b_n(t)$ to be of order $1/\epsilon_n$. Then the probability

$$P\left(\lambda_1 > M_1^{a_n(s)} \right) \sim P\left(\lambda_1 > 1/\epsilon_n \right) = P\left(\lambda_1^2 > 1/\epsilon_n^2 \right)$$

which is $o(\epsilon_n^2)$ since $E(\lambda_1^2) < \infty$. Similarly, when $E(\lambda_2^2) < \infty$, $1/\epsilon_n^2 P\left(\mu_2 > \mu_2^{b_n(t)} \right) \to 0$. Therefore, Assumption 3.3.2 is satisfied. $\square$

3.5 Proofs

3.5.1 Proof of Consistency

Here are some simplified notations in the proof: $F_1(k_1) = F_1(k_1|X_1)$, $F_2(k_2) = F(k_2|X_2)$, $H(s,t) = H(s,t;\mathbf{X})$, $H(s,t;\theta) = H(s,t;\mathbf{X},\theta)$, and $\hat{C}(s,t) = \hat{C}(s,t;\beta)$, where β is the

underlying parameter.

Proof of Lemma 3.3.1. Recall that $\epsilon_n \to 0$. For $v_1^k, k = 1, 2, \ldots$ as in Section 3.3, taking minimum for the first n elements, $u_n = \min_{k=1,\ldots,n} v_1^k$ is a nonzero decreasing sequence. Therefore, an appropriate order of $a_n(s)$ can be chosen such that $u_{a_n(s)} > \epsilon_n$, i.e., $v_1^k > \epsilon_n$ for $k \leq a_n(s)$. □

To show the asymptotic properties of $\hat{C}(s,t)$ defined in (3.5), we analyze it by pieces. We first show the denominator is a consistent estimator of $f_{H(s,t)}(s,t)$. Then, we show the consistency of the numerator.

Denote the denominator as

$$\hat{f}_{H(s,t)}(s,t) = \frac{1}{n\epsilon_n^2} \sum_{i=1}^{n} K\left[\epsilon_n^{-1}(H_{i1}(s) - s), \epsilon_n^{-1}(H_{i2}(t) - t)\right].$$

Lemma 3.5.1 shows the consistency of the denominator.

Lemma 3.5.1. *Under Assumptions 3.3.1 and 3.3.2,*

$$\hat{f}_{H(s,t)}(s,t) \to_p f_{H(s,t)}(s,t). \tag{3.16}$$

Proof. Recall that K is a bounded on compact support. Without loss of generality, in the proof we assume $K(u,v) \leq 1$ with support $|(u,v)| \leq 1$.

Let $\hat{f}_{H(s,t)}(s,t) = \frac{1}{n} \sum_{i=1}^{n} T_{ni}$, where $T_{ni} = 1/\epsilon_n^2 K\left[\epsilon_n^{-1}(H_{i1}(s) - s), \epsilon_n^{-1}(H_{i2}(t) - t)\right]$. That is, $\hat{f}_{H(s,t)}(s,t)$ is the summation of a triangular array. We demonstrate the consistency of $\hat{f}_{H(s,t)}(s,t)$ using the weak law of large numbers (WLLN) for triangular arrays. It is sufficient to show

$$\frac{1}{n} \sum_{i=1}^{n} \mathrm{E}T_{ni} \to f_{H(s,t)}(s,t), \tag{3.17}$$

$$\frac{1}{n}\mathrm{E}T_{ni}^2 \to 0. \tag{3.18}$$

First, to show (3.17), we divide the range of μ_1 and μ_2 into four pieces, i.e.

$$
\begin{aligned}
\mathrm{E}T_{ni} =& \mathrm{E}\left[T_{ni}1\left(\mu_1 \le M_1^{a_n(s)}, \mu_2 \le M_2^{b_n(t)}\right)\right] + \mathrm{E}\left[T_{ni}1\left(\mu_1 \le M_1^{a_n(s)}, \mu_2 > M_2^{b_n(t)}\right)\right] \\
&+ \mathrm{E}\left[T_{ni}1\left(\mu_1 > M_1^{a_n(s)}, \mu_2 \le M_2^{b_n(t)}\right)\right] + \mathrm{E}\left[T_{ni}1\left(\mu_1 > M_1^{a_n(s)}, \mu_2 > M_2^{b_n(t)}\right)\right] \\
:=& T_1 + T_2 + T_3 + T_4.
\end{aligned}
$$

$$(3.19)$$

We analyze the four pieces one by one.

Let f_{μ_1,μ_2} denote the joint density of (μ_1, μ_2). The first term of (3.19) equals

$$
T_1 = \sum_{k_1=0}^{a_n(s)} \sum_{k_2=0}^{b_n(t)} T_1(k_1, k_2),
$$

where

$$
T_1(k_1, k_2) = \frac{1}{\epsilon_n^2} \int_{M_1^{k_1}}^{M_1^{k_1+1}} \int_{M_2^{k_2}}^{M_2^{k_2+1}} K\left[\epsilon_n^{-1}(F_1(k_1) - s), \epsilon_n^{-1}(F_2(k_2) - t)\right] f_{\mu_1,\mu_2}(\mu_1, \mu_2)\mathrm{d}\mu_1\mathrm{d}\mu_2.
$$

$$(3.20)$$

Recall that K takes nonzero values only when

$$
|(F_1(k_1) - s, F_2(k_2) - t)| \le \epsilon_n. \tag{3.21}
$$

For $k_1 \le a_n(s), k_2 \le b_n(t)$, a necessary condition for (3.21) to hold is that $M_1^{k_1} \le \mu_1 < M_1^{k_1+1}, M_2^{k_2} \le \mu_2 < M_2^{k_2+1}$. Therefore, we can write the limits in (3.20) as $(-\infty, \infty)$. That is,

$$
\begin{aligned}
T_1(k_1, k_2) =& \frac{1}{\epsilon_n^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K\left[\epsilon_n^{-1}(F_1(k_1) - s), \epsilon_n^{-1}(F_2(k_2) - t)\right] f_{\mu_1,\mu_2}(\mu_1, \mu_2)\mathrm{d}\mu_1\mathrm{d}\mu_2 \\
=& \frac{1}{\epsilon_n^2} \int K\left(\frac{a - s}{\epsilon_n}, \frac{b - t}{\epsilon_n}\right) f_{F_1(k_1), F_2(k_2)}(a, b)\mathrm{d}a\mathrm{d}b \\
=& \int K(u, v) f_{F_1(k_1), F_2(k_2)}(s + u\epsilon_n, t + v\epsilon_n)\mathrm{d}u\mathrm{d}v,
\end{aligned}
$$

where the last equation is derived by substitution. A Taylor series expansion up to first order for $f_{F_1(k_1),F_2(k_2)}$ yields

$$f_{F_1(k_1),F_2(k_2)}(s + u\epsilon_n, t + v\epsilon_n)$$
$$= f_{F_1(k_1),F_2(k_2)}(s,t) + f_{F_1(k_1),F_2(k_2),1}(s,t)u\epsilon_n + f_{F_1(k_1),F_2(k_2),2}(s,t)v\epsilon_n +$$
$$\frac{1}{2}f_{F_1(k_1),F_2(k_2),11}(s,t)u^2\epsilon_n^2 + \frac{1}{2}f_{F_1(k_1),F_2(k_2),22}(s,t)v^2\epsilon_n^2 + f_{F_1(k_1),F_2(k_2),12}(s,t)uv\epsilon_n^2 + o(\epsilon_n^2).$$

Since K is symmetric, $\int K(u,v)u\,du\,dv = 0$, and $\int K(u,v)uv\,du\,dv = 0$. Moreover, $\int K(u,v)\,du\,dv = 1$. Denote $f_{H(s,t),n}(s,t) = \sum_{k_1=0}^{a_n(s)}\sum_{k_2=0}^{b_n(t)} f_{F_1(k_1),F_2(k_2)}(s,t)$, then we have

$$T_1 = f_{H(s,t),n}(s,t) + o(\epsilon_n),$$

where the residual term is $o(\epsilon_n)$ since $\sum_{k_1=0}^{a_n(s)}\sum_{k_2=0}^{b_n(t)} f_{F_1(k_1),F_2(k_2),jj}(s,t) < \infty, j = 1,2$, where $f_{F_1(k_1),F_2(k_2),jj}$ is the second order derivatives of $f_{F_1(k_1),F_2(k_2)}$ with respect to jth component, by Assumption 3.3.1. As $a_n(s)$ and $b_n(t)$ go to infinity with n,

$$T_1 \to f_{H(s,t)}(s,t).$$

Then, we consider the second term T_2. As $|K(u,v)| \le 1$,

$$T_2 \le \frac{1}{\epsilon_n^2}\int_{-\infty}^{\infty}\int_{M_2^{b_n(t)}}^{\infty} f_{\mu_1,\mu_2}(\mu_1,\mu_2)\,d\mu_1 d\mu_2 = \frac{1}{\epsilon_n^2}P\left(\mu_2 > M_2^{b_n(t)}\right). \tag{3.22}$$

By Assumption 3.3.2, $T_2 \to 0$. Similar arguments can be used to show that

$$T_3 \le \frac{1}{\epsilon_n^2}P\left(\mu_1 > M_1^{a_n(s)}\right) \to 0,$$
$$T_4 \le \frac{1}{\epsilon_n^2}P\left(\mu_1 > M_1^{a_n(s)}, \mu_2 > M_2^{b_n(s)}\right) \to 0.$$

So (3.17) follows immediately.

Finally, we show (3.18). Since K is bounded by 1,

$$T_{ni}^2 \leq \frac{1}{\epsilon_n^2} T_{ni}.$$

As T_{ni} is positive, combining with (3.17) and that $n\epsilon_n^2 \to \infty$, we have

$$\frac{1}{n} \mathrm{E} T_{ni}^2 \leq \frac{1}{n\epsilon_n^2} \mathrm{E} T_{ni} \to 0.$$

The lemma follows the WLLN for triangular arrays. $\qquad\square$

Proof of Theorem 3.3.1. Given Lemma 3.5.1, it is sufficient to show that the numerator of (3.5) is a consistent estimator of $f_{H(s,t)}(s,t)C(s,t)$. Similar to the denominator, the numerator of (3.5), denoted as

$$\hat{N}(s,t) = \frac{1}{n\epsilon_n^2} \sum_{i=1}^{n} K \left[\epsilon_n^{-1}(H_{i1}(s) - s), \epsilon_n^{-1}(H_{i2}(t) - t) \right] Y_i(\beta) \qquad (3.23)$$

is the summation of a triangular array, i.e.,

$$\hat{N}(s,t) = \frac{1}{n} \sum_{i=1}^{n} V_{ni},$$

where $V_{ni} = 1/\epsilon_n^2 K \left[\epsilon_n^{-1}(H_{i1}(s) - s), \epsilon_n^{-1}(H_{i2}(t) - t) \right] Y_i(\beta)$. It is sufficient to show

$$\frac{1}{n} \sum_{i=1}^{n} \mathrm{E} V_{ni} \to f_{H(s,t)}(s,t)C(s,t), \qquad (3.24)$$

$$\frac{1}{n} \mathrm{E} V_{ni}^2 \to 0. \qquad (3.25)$$

We first show (3.24). Note that

$$\mathrm{E}\left[V_{ni} | H_{i1}(s), H_{i2}(t) \right] = K \left[\epsilon_n^{-1}(H_{i1}(s) - s), \epsilon_n^{-1}(H_{i2}(t) - t) \right] C \left[H_{i1}(s), H_{i2}(t) \right].$$

Similar ideas used to show (3.17) lead to the approximation

$$\mathrm{E}\left(V_{ni}\right) \approx \sum_{k_1=0}^{a_n(s)} \sum_{k_2=0}^{b_n(t)} \int K(u,v) f_{F_1(k_1),F_2(k_2)}(s+u\epsilon_n, t+v\epsilon_n) C(s+u\epsilon_n, t+v\epsilon_n) \mathrm{d}u \mathrm{d}v.$$

Taking the product of Taylor expansions of $f_{F_1(k_1),F_2(k_2)}$ and C at (s,t) yields

$$\mathrm{E}\left(V_{ni}\right) \approx \sum_{k_1=0}^{a_n(s)} \sum_{k_2=0}^{b_n(t)} C(s,t) f_{F_1(k_1),F_2(k_2)}(s,t) + o(\epsilon_n).$$

Note that some terms are eliminated due to the symmetry of K. When n approaches infinity, (3.24) follows immediately. In addition, since K and $Y_i(\beta)$ are bounded by 1,

$$\frac{1}{n}\mathrm{E}\left(V_{ni}^2\right) \leq \frac{1}{n\epsilon_n^2}\mathrm{E}\left(V_{ni}\right) \to 0.$$

So (3.25) holds and the stated result follows WLLN for triangular arrays. $\quad\square$

3.5.2 Proof of Theorem 3.3.2

Proof of Theorem 3.3.2. Given consistency of the denominator in Section 3.5.1, now we are in a position to show the weak convergence of the numerator. We check the bias and variance of the numerator and show they both converge to 0 at the appropriate rate.

Comparing $\hat{C}(s,t)$ and $C(s,t)$ yields

$$\hat{C}(s,t) = C(s,t) + \frac{\hat{m}_1(s,t)}{\hat{f}_{H(s,t)}(s,t)} + \frac{\hat{m}_2(s,t)}{\hat{f}_{H(s,t)}(s,t)},$$

where

$$\hat{m}_1(s,t) = \frac{1}{n\epsilon_n^2} \sum_{i=1}^{n} K\left[\epsilon_n^{-1}\left(H_{i1}(s) - s\right), \epsilon_n^{-1}(H_{i2}(t) - t)\right] \left[C\left[H_{i1}(s), H_{i2}(t)\right] - C(s,t)\right],$$

$$\hat{m}_2(s,t) = \frac{1}{n\epsilon_n^2} \sum_{i=1}^{n} K\left[\epsilon_n^{-1}\left(H_{i1}(s) - s\right), \epsilon_n^{-1}(H_{i2}(t) - t)\right] \left\{Y_i(\beta) - C\left[H_{i1}(s), H_{i2}(t)\right]\right\}.$$

Among them, $\hat{m}_1(s,t)$ contributes to the bias while $\hat{m}_2(s,t)$ contributes the variance of $\hat{C}(s,t)$.

Variance. Since $\mathrm{E}\left(Y_i(\beta)|\mu_{i1},\mu_{i2}\right) = C\left[H_{i1}(s),H_{i2}(t)\right]$, one has $\mathrm{E}(\hat{m}_2(s,t)|\mu_{i1},\mu_{i2}) = 0$, which leads to that $\mathrm{E}\left[\hat{m}_2(s,t)\right] = 0$. Thus, we focus on the variance of $\hat{m}_2(s,t)$. Note that

$$\mathrm{Var}\left[\hat{m}_2(s,t)\right] = \frac{1}{n\epsilon_n^4}\mathrm{E}\left[K\left[\epsilon_n^{-1}\left(H_{i1}(s)-s\right),\epsilon_n^{-1}(H_{i2}(t)-t)\right]\left\{Y_i(\beta)-C\left[H_{i1}(s),H_{i2}(t)\right]\right\}\right]^2$$

(3.26)

To compute the variance of $\hat{m}_2(s,t)$, we condition on $H_i(s,t)$

$$\mathrm{E}\left(\left\{Y_i(\beta)-C\left[H_{i1}(s),H_{i2}(t)\right]\right\}^2|H_{i1}(s)=a,H_{i2}(t)=b\right) = C(a,b)\left[1-C(a,b)\right] := \sigma^2(a,b).$$

Since K and σ are bounded, arguments analogous to those use to prove (3.17) lead to the approximation

$$\mathrm{Var}\left[\hat{m}_2(s,t)\right]$$

$$\approx \frac{1}{n\epsilon_n^4}\sum_{k_1=0}^{a_n(s)}\sum_{k_2=0}^{b_n(t)}\int_{M_1^{k_1}}^{M_1^{k_1+1}}\int_{M_2^{k_2}}^{M_2^{k_2+1}}K\left[\epsilon_n^{-1}(F_1(k_1)-s),\epsilon_n^{-1}(F_2(k_2)-t)\right]^2\sigma^2\left[F_1(k_1),F_2(k_2)\right]\cdot$$

$$f_{\mu_1,\mu_2}(\mu_1,\mu_2)\mathrm{d}\mu_1\mathrm{d}\mu_2 + o\left(\frac{1}{n\epsilon_n^2}\right)$$

$$= \frac{1}{n\epsilon_n^4}\sum_{k_1=0}^{a_n(s)}\sum_{k_2=0}^{b_n(t)}\int K\left(\frac{a-s}{\epsilon_n},\frac{b-t}{\epsilon_n}\right)^2\sigma^2(a,b)f_{F_1(k_1),F_2(k_2)}(a,b)\mathrm{d}a\mathrm{d}b + o\left(\frac{1}{n\epsilon_n^2}\right)$$

$$= \frac{1}{n\epsilon_n^2}\sum_{k_1=0}^{a_n(s)}\sum_{k_2=0}^{b_n(t)}\int K(u,v)^2\sigma^2(s+u\epsilon_n,t+v\epsilon_n)f_{F_1(k_1),F_2(k_2)}(s+u\epsilon_n,t+v\epsilon_n)\mathrm{d}u\mathrm{d}v + o\left(\frac{1}{n\epsilon_n^2}\right).$$

As C and $f_{F_1(k_1),F_2(k_2)}$ are twice continuously differentiable over V, $\sigma^2 f_{F_1(k_1),F_2(k_2)}$ carries over the differentiability as a function of them. Denote $\alpha_{21}(s,t) =$

$R_2(K)\sigma^2(s,t)f_{H(s,t),n}(s,t)$. The Taylor series expansion of $\sigma^2 f_{F_1(k_1),F_2(k_2)}$ yields

$$\text{Var}\left[\hat{m}_2(s,t)\right] = \frac{1}{n\epsilon_n^2}\alpha_{21}(s,t) + o\left(\frac{1}{n\epsilon_n^2}\right).$$

Note that the term $\hat{m}_2(s,t)$ is the summation of a triangular array. To establish the asymptotic distribution of $\hat{m}_2(s,t)$, we now verify that Lyapunov's central limit theorem holds. Denote

$$w_{ni}(s,t) = \frac{1}{n\epsilon_n^2}K\left[\epsilon_n^{-1}\left(H_{i1}(s)-s\right), \epsilon_n^{-1}(H_{i2}(t)-t)\right]\left\{Y_i(\beta) - C\left[H_{i1}(s), H_{i2}(t)\right]\right\},$$

then $\hat{m}_2(s,t) = \sum_{i=1}^{n} w_{ni}$. It is sufficient to show

$$\frac{\sum_{i=1}^{n}\text{E}\left|w_{ni}(s,t)\right|^3}{\left(\sum_{i=1}^{n}\text{Var}\left[w_{ni}(s,t)\right]\right)^{3/2}} \to 0. \tag{3.27}$$

Note that

$$\text{Var}\left[w_{ni}(s,t)\right] = \frac{1}{n^2\epsilon_n^2}\alpha_{21}(s,t) + o\left(\frac{1}{n^2\epsilon_n^2}\right),$$

$$\text{E}\left|w_{ni}(s,t)\right|^3 \leq \frac{f_{H(s,t),n}(s,t)}{n^3\epsilon_n^4}\int K(u,v)^3 \mathrm{d}u\mathrm{d}v + o\left(\frac{1}{n^3\epsilon_n^4}\right) = O\left(\frac{1}{n^3\epsilon_n^4}\right).$$

Hence, (3.27) follows immediately, and the Lyapunov condition is satisfied. Therefore,

$$\sqrt{n\epsilon_n^2}\frac{\hat{m}_2(s,t)}{\sqrt{\alpha_{21}(s,t)}} \to_d N(0,1).$$

Since $f_{H(s,t),n}(s,t) \to f_{H(s,t)}(s,t)$,

$$\sqrt{n\epsilon_n^2}\hat{m}_2(s,t) \to_d N\left(0, R_2(K)C(s,t)\left[1-C(s,t)\right]f_{H(s,t)}(s,t)\right).$$

By Slutsky's theorem, substitution of $\hat{f}_{H(s,t)}(s,t)$ by $f_{H(s,t)}(s,t)$ yields

$$\sqrt{n\epsilon_n^2}\,\frac{\hat{m}_2(s,t)}{\hat{f}_{H(s,t)}(s,t)} \to_d N\left(0,\frac{C(s,t)\left[1-C(s,t)\right]R_2(K)}{f_{H(s,t)}(s,t)}\right). \tag{3.28}$$

Bias. The mean of $\hat{m}_1(s,t)$ is

$$\mathrm{E}\left[\hat{m}_1(s,t)\right] = \frac{1}{\epsilon_n^2}\mathrm{E}\left\{K\left[\epsilon_n^{-1}\left(H_{i1}(s)-s\right),\epsilon_n^{-1}(H_{i2}(t)-t)\right]\left[C\left[H_{i1}(s),H_{i2}(t)\right]-C(s,t)\right]\right\}.$$

Since C satisfies Lipschitz condition as in Assumption 3.3.4, given $|(H_{i1}(s)-s,H_{i2}(t)-t)| < \epsilon_n$, under which $K\left[\epsilon_n^{-1}\left(H_{i1}(s)-s\right),\epsilon_n^{-1}(H_{i2}(t)-t)\right]$ is nonzero,

$$|C\left[H_{i1}(s),H_{i2}(t)\right]-C(s,t)| \le \alpha_1\epsilon_n^2.$$

Similar arguments used to show (3.17) lead to the approximation

$$\mathrm{E}\left[\hat{m}_1(s,t)\right] \approx \frac{1}{\epsilon_n^2}\sum_{k_1=0}^{a_n(s)}\sum_{k_2=0}^{b_n(t)}\int K\left(\frac{a-s}{\epsilon_n},\frac{b-t}{\epsilon_n}\right)\left[C(a,b)-C(s,t)\right]f_{F_1(k_1),F_2(k_2)}(a,b)\mathrm{d}a\mathrm{d}b + o(\epsilon_n^2)$$

$$= \sum_{k_1=0}^{a_n(s)}\sum_{k_2=0}^{b_n(t)}\int K(u,v)\left[C(s+u\epsilon_n,t+v\epsilon_n)-C(s,t)\right]\cdot$$

$$f_{F_1(k_1),F_2(k_2)}(s,t)(s+u\epsilon_n,t+v\epsilon_n)\mathrm{d}u\mathrm{d}v + o(\epsilon_n^2).$$

Taking the product of the Taylor series expansions for C up to second order and first order for $f_{F_1(k_1),F_2(k_2)}$ at (s,t), one gets

$$\mathrm{E}\left[\hat{m}_1(s,t)\right] = \sum_{k_1=0}^{a_n(s)}\sum_{k_2=0}^{b_n(t)}\int\left\{K(u,v)\epsilon_n^2\left[\frac{1}{2}C_{11}(s,t)f_{F_1(k_1),F_2(k_2)}(s,t)u^2+\right.\right.$$

$$\frac{1}{2}C_{22}(s,t)f_{F_1(k_1),F_2(k_2)}(s,t)v^2 + C_1(s,t)f_{F_1(k_1),F_2(k_2),1}(s,t)u^2+$$

$$\left.\left.C_2(s,t)f_{F_1(k_1),F_2(k_2),2}(s,t)v^2\right]\right\}\mathrm{d}u\mathrm{d}v + o(\epsilon_n^2).$$

Let $f_{H(s,t),n,j}(s,t), j = 1,2$ denote the partial derivatives of $f_{H(s,t),n}(s,t)$ and define

$$\zeta_n(s,t) = \frac{1}{2}C_{11}(s,t) + \frac{1}{2}C_{22}(s,t) + \frac{C_1(s,t)f_{H(s,t),n,1}(s,t)}{f_{H(s,t),n}(s,t)} + \frac{C_2(s,t)f_{H(s,t),n,2}(s,t)}{f_{H(s,t),n}(s,t)}.$$

Recall the notation $\kappa_2 = \int u^2 K(u,v)\mathrm{d}u$, hence we obtain

$$\mathrm{E}\left[\hat{m}_1(s,t)\right] = \kappa_2 \zeta_n(s,t) f_{H(s,t),n}(s,t)\epsilon_n^2 + o(\epsilon_n^2). \tag{3.29}$$

We now compute the variance of $\hat{m}_1(s,t)$

$$\mathrm{Var}\left[\hat{m}_1(s,t)\right]$$

$$= \frac{1}{n\epsilon_n^4}\mathrm{Var}\left\{K\left[\epsilon_n^{-1}\left(H_{i1}(s) - s\right), \epsilon_n^{-1}(H_{i2}(t) - t)\right]\left[C\left[H_{i1}(s), H_{i2}(t)\right] - C(s,t)\right]\right\}$$

$$= \frac{1}{n\epsilon_n^4}\left(\mathrm{E}\left\{K\left[\epsilon_n^{-1}\left(H_{i1}(s) - s\right), \epsilon_n^{-1}(H_{i2}(t) - t)\right]\left[C\left[H_{i1}(s), H_{i2}(t)\right] - C(s,t)\right]\right\}^2\right.$$

$$\left. - \left\{\mathrm{E}\left(K\left[\epsilon_n^{-1}\left(H_{i1}(s) - s\right), \epsilon_n^{-1}(H_{i2}(t) - t)\right]\left[C\left[H_{i1}(s), H_{i2}(t)\right] - C(s,t)\right]\right)\right\}^2\right). \tag{3.30}$$

The first term of (3.30) is

$$\frac{1}{n\epsilon_n^4}\mathrm{E}\left\{K\left[\epsilon_n^{-1}\left(H_{i1}(s) - s\right), \epsilon_n^{-1}(H_{i2}(t) - t)\right]\left[C\left[H_{i1}(s), H_{i2}(t)\right] - C(s,t)\right]\right\}^2$$

$$= \frac{1}{n\epsilon_n^4}\mathrm{E}\left(\sum_{k_1=0}^{a_n(s)}\sum_{k_2=0}^{b_n(t)}K\left(\frac{F_{i1}(k_1) - s}{\epsilon_n}, \frac{F_{i2}(k_2) - t}{\epsilon_n}\right)\left[C\left[F_{i1}(k_1), F_{i2}(k_2)\right] - C(s,t)\right]\right)^2 + o\left(\frac{1}{n}\right)$$

$$= \frac{1}{n\epsilon_n^2}\sum_{k_1=0}^{a_n(s)}\sum_{k_2=0}^{b_n(t)}\int K(u,v)^2\left[C(s + u\epsilon_n, t + v\epsilon_n) - C(s,t)\right]^2 f_{F_1(k_1),F_2(k_2)}(s + u\epsilon_n, t + v\epsilon_n)\mathrm{d}u\mathrm{d}v$$

$$+ o\left(\frac{1}{n}\right),$$

where the residuals are obtained by similar techniques used to compute $\mathrm{E}\left[\hat{m}_1(s,t)\right]$. By Taylor expansions, we have $\left[C(s + u\epsilon_n, t + v\epsilon_n) - C(s,t)\right]^2 f_{F_1(k_1),F_2(k_2)}(s + u\epsilon_n, t + v\epsilon_n) = O(\epsilon_n^2)$. Thus, the first term of (3.30) is of order $O(1/n)$.

Then, we check the second term of (3.30) . From Equation (3.29),

$$\frac{1}{n\epsilon_n^4} \left\{ \mathrm{E} \left(K \left[\epsilon_n^{-1} \left(H_{i1}(s) - s \right), \epsilon_n^{-1}(H_{i2}(t) - t) \right] \left[C \left[H_{i1}(s), H_{i2}(t) \right] - C(s,t) \right] \right) \right\}^2 = O\left(\frac{1}{n}\right).$$

(3.31)

Therefore,

$$\mathrm{Var}\left(\sqrt{n\epsilon_n^2} \hat{m}_1(s,t) \right) = O(\epsilon_n^2) \to 0. \tag{3.32}$$

Equations (3.29) and (3.32) entail that

$$\sqrt{n\epsilon_n^2} \left(\hat{m}_1(s,t) - \kappa_2 \zeta_n(s,t) f_{H(s,t),n}(s,t)\epsilon_n^2 \right) \to_p 0.$$

Recall $\zeta(s,t)$ defined in (3.8), it can be easily seen that $\zeta_n(s,t) \to \zeta(s,t)$. As $a_n(s)$ and $b_n(t)$ go to infinity, we have

$$\sqrt{n\epsilon_n^2} \left(\hat{m}_1(s,t) - \kappa_2 \zeta(s,t) f_{H(s,t)}(s,t)\epsilon_n^2 \right) \to_p 0.$$

Together with (3.16), it follows

$$\sqrt{n\epsilon_n^2} \left(\frac{\hat{m}_1(s,t)}{\hat{f}_{H(s,t)}(s,t)} - \kappa_2 \frac{f_{H(s,t)}(s,t)}{\hat{f}_{H(s,t)}(s,t)} \zeta(s,t)\epsilon_n^2 \right) \to_p 0.$$

Recall $n\epsilon_n^6 = O(1)$, hence we also have

$$\sqrt{n\epsilon_n^2} \left(\kappa_2 \frac{f_{H(s,t)}(s,t)}{\hat{f}_{H(s,t)}(s,t)} \zeta(s,t)\epsilon_n^2 - \kappa_2 \zeta(s,t)\epsilon_n^2 \right) \to_p 0. \tag{3.33}$$

Therefore,

$$\sqrt{n\epsilon_n^2} \left(\frac{\hat{m}_1(s,t)}{\hat{f}_{H(s,t)}(s,t)} - \kappa_2 \zeta(s,t)\epsilon_n^2 \right) \to_p 0. \tag{3.34}$$

Summing up (3.28) and (3.34) finishes the proof.

$\square$

Hence the AMSE of $\hat{C}(s,t)$ is $(C(s,t)\left[1-C(s,t)\right]R_2(K))/(n\epsilon_n^2 f_{H(s,t)}(s,t))$ + $\kappa_2^2\zeta(s,t)^2\epsilon_n^4$.

The above theorems guarantee the identifiability of underlying copula. With Assumptions 3.3.1, 3.3.2, 3.3.3, and 3.3.4, if there exists another copula $\tilde{C}$ compatible with data, the pointwise difference between C and $\tilde{C}$ at $(s,t)\in V$ is

$$\left|C(s,t)-\tilde{C}(s,t)\right|^2 \leq \mathrm{E}\left|\hat{C}(s,t)-C(s,t)\right|^2 + \mathrm{E}\left|\hat{C}(s,t)-\tilde{C}(s,t)\right|^2 \to 0.$$

Since $\tilde{C}(s,t)-C(s,t)$ does not change with n, it has to be that $\tilde{C}(s,t)-C(s,t)=0$ for any $(s,t)\in V$. That is, the copula is identifiable at V.

3.5.3   Proof of Theorem 3.3.3

We now demonstrate the asymptotic properties of the copula estimator defined in (3.5) when the marginal parameters are unknown and the estimates are plugged in. We first analyze the numerator and denominator of the estimator separately in Lemmas 3.5.2 and 3.5.4. Finally, the distribution of the copula estimator follows as stated in Theorem 3.3.3.

For the following results Lemmas 3.5.2, 3.5.3, and 3.5.4, we prove them with uniform kernel. The proof for other compacted supported kernels is a trivial extension from the uniform kernel with Lipschitz conditions of the kernel function. Let $A_n(s,t)$ denote the neighborhood of (s,t) with radius ϵ_n.

Lemma 3.5.2 (Consistency of Denominator). *Under Assumptions 3.3.1, 3.3.2, 3.3.5, and 3.3.6,*

$$\frac{1}{n\epsilon_n^2}\sum_{i=1}^{n}\left\{1\left(H_i(s,t;\hat{\beta})\in A_n(s,t)\right)-1\left[H_i(s,t)\in A_n(s,t)\right]\right\}\to_p 0. \qquad (3.35)$$

Therefore,

$$\frac{1}{n\epsilon_n^2} \sum_{i=1}^n K\left[\epsilon_n^{-1}(H_{i1}(s;\hat{\beta}) - s), \epsilon_n^{-1}(H_{i2}(s;\hat{\beta}) - t)\right] \to_p f_{H(s,t)}(s,t). \qquad (3.36)$$

Proof. Different aspects of the method of proof can be found in Sukhatme (1958), Randles (1984), and Frees (1995b). Denote

$$l_{ni}(s,t;\theta) = 1\left(H_i(s,t;\theta) \in A_n(s,t)\right)$$

and

$$S_n(\theta; s,t) = \frac{1}{n\epsilon_n^2} \sum_{i=1}^n \left[l_{ni}(s,t;\beta) - l_{ni}(s,t;\theta)\right].$$

For arbitrary $\epsilon > 0$, for any $\xi > 0$, let γ_ξ be the constant in Assumption 3.3.6. To show (3.35), we calculate the probability

$$
\begin{aligned}
P\left(\left|S_n(\hat{\beta}; s,t)\right| > \epsilon\right) &= P\left(\left|S_n(\hat{\beta}; s,t)\right| > \epsilon, \hat{\beta} \in B(\beta, n^{-1/2}\gamma_\xi)\right) \\
&\quad + P\left(\left|S_n(\hat{\beta}; s,t)\right| > \epsilon, \hat{\beta} \notin B(\beta, n^{-1/2}\gamma_\xi)\right) \\
&\leq P\left(\sup_{\theta \in B(\beta, n^{-1/2}\gamma_\xi)} |S_n(\theta; s,t)| > \epsilon\right) + P\left(\hat{\beta} \notin B(\beta, n^{-1/2}\gamma_\xi)\right) \\
&:= M_1 + M_2.
\end{aligned}
$$

By Assumption 3.3.6,

$$M_2 < \xi/2.$$

We now check M_1. Since $|l_{ni}(s,t;\beta) - l_{ni}(s,t;\theta)| \leq 1$, using a similar method as in Frees (1995a), we can see there are two cases $|l_{ni}(s,t;\beta) - l_{ni}(s,t;\theta)|$ can be 1, i.e.,

$$|l_{ni}(s,t;\beta) - l_{ni}(s,t;\theta)| \leq \begin{cases} 1(|H_i(s,t;\theta) - (s,t)| > \epsilon_n, |H_i(s,t) - (s,t)| \leq \epsilon_n) := J_1, \\ 1(|H_i(s,t;\theta) - (s,t)| \leq \epsilon_n, |H_i(s,t) - (s,t)| > \epsilon_n) := J_2. \end{cases}$$

For the first case J_1,

$$J_1 \leq 1(\epsilon_n < |H_i(s,t;\theta) - (s,t)|).$$

Further, subtracting $|H_i(s,t) - (s,t)|$ from both sides yields

$$J_1 \leq 1\left(\epsilon_n - |H_i(s,t) - (s,t)| < |H_i(s,t;\theta) - (s,t)| - |H_i(s,t) - (s,t)|\right).$$

Similarly, for the second case J_2

$$J_2 \leq 1\left(\epsilon_n \geq |H_i(s,t;\theta) - (s,t)|\right).$$

So we have

$$J_2 \leq 1\left(|H_i(s,t) - (s,t)| - \epsilon_n \leq |H_i(s,t) - (s,t)| - |H_i(s,t;\theta) - (s,t)|\right).$$

Since if $J_1 = 1$, $|H_i(s,t) - (s,t)| \leq \epsilon_n$, and when $J_2 = 1$, $|H_i(s,t) - (s,t)| > \epsilon_n$, we can summarize these two cases obtaining

$$|l_{ni}(s,t;\beta) - l_{ni}(s,t;\theta)| \leq 1\left(||H_i(s,t) - (s,t)| - \epsilon_n| \leq ||H_i(s,t) - (s,t)| - |H_i(s,t;\theta) - (s,t)||\right)$$

$$\leq 1\left(||H_i(s,t) - (s,t)| - \epsilon_n| \leq \alpha_2 |\beta - \theta|\right),$$

(3.37)

where the second inequality is due to Assumption 3.3.5.

Now we take supremum with respect to θ over $B(\beta, n^{-1/2}\gamma_\xi)$, defining

$$\eta_n(s,t;X_i) = \sup_{\theta \in B(\beta,n^{-1/2}\gamma_\xi)} |l_{ni}(s,t;\theta) - l_{ni}(s,t;\beta)|.$$

From (3.37), there exists a constant α_2 such that

$$\eta_n(s,t;X_i) \leq 1\left(||H_i(s,t) - (s,t)| - \epsilon_n| \leq \alpha_2 n^{-1/2}\gamma_\xi\right).$$

By Assumption 3.3.1, the density of $H_i(s,t)$ is bounded. Thus $|H_i(s,t) - (s,t)| - \epsilon_n$ which is the linear transformation result from $H_i(s,t)$ also has a bounded density. For all n, i, and (s,t), there exists a constant α_3 such that

$$\mathrm{E}\eta_n(s,t;X_i) \leq \alpha_3\gamma_\xi^2 n^{-1}. \tag{3.38}$$

Note that $\sup_{\theta \in B(\beta, n^{-1/2}\gamma_\xi)} |S_n(\theta; s,t)| \leq 1/(n\epsilon_n^2)\sum_{i=1}^n \eta_n(s,t;X_i)$. Therefore,

$$M_1 \leq P\left(\frac{1}{n\epsilon_n^2}\sum_{i=1}^n \eta_n(s,t;X_i) > \epsilon\right)$$

$$\leq P\left(\left(\frac{1}{n\epsilon_n^2}\sum_{i=1}^n [\eta_n(s,t;X_i) - \mathrm{E}\eta_n(s,t;X_i)] + \frac{1}{n\epsilon_n^2}\sum_{i=1}^n \mathrm{E}\eta_n(s,t;X_i)\right) > \epsilon\right)$$

From (3.38),

$$\frac{1}{n\epsilon_n^2}\sum_{i=1}^n \mathrm{E}\eta_n(s,t;X_i) \leq \frac{\alpha_3\gamma_\xi^2}{n\epsilon_n^2} \to 0.$$

Hence, when n gets large,

$$\frac{1}{n\epsilon_n^2}\sum_{i=1}^n \mathrm{E}\eta_n(s,t;X_i) < \epsilon/2.$$

By applying the Chebyshev's inequality,

$$M_1 \leq P\left(\frac{1}{n\epsilon_n^2}\sum_{i=1}^n [\eta_n(s,t;X_i) - \mathrm{E}\eta_n(s,t;X_i)] > \epsilon/2\right)$$

$$\leq \frac{1}{(\epsilon/2)^2}\mathrm{Var}\left(\frac{1}{n\epsilon_n^2}\sum_{i=1}^n [\eta_n(s,t;X_i) - \mathrm{E}\eta_n(s,t;X_i)]\right).$$

Note that $\eta_n(s,t;X_i)^2 = \eta_n(s,t;X_i)$. By applying (3.38) we have

$$M_1 \leq \frac{1}{(\epsilon/2)^2}\frac{1}{(n\epsilon_n^2)^2}\sum_{i=1}^n \left[\mathrm{E}\eta_n(s,t;X_i)^2 - [\mathrm{E}\eta_n(s,t;X_i)]^2\right] \leq \frac{1}{(\epsilon/2)^2(n\epsilon_n^2)^2}\alpha_3\gamma_\xi^2.$$

Therefore, when n is large enough $M_1 < \xi/2$. Now (3.35) follows from the fact that for arbitrary ϵ and $\xi > 0$,

$$P\left(\left|S_n(\hat{\beta}; s,t)\right| > \epsilon\right) < \xi.$$

Finally, note that (3.36) follows from (3.16) and (3.35), and the proof is finished. □

Define

$$h_{ni}(s,t;\theta) = 1\left(H_i(s,t;\theta) \in A_n(s,t)\right) Y_i(\theta),$$

where $Y_i(\theta)$ is defined in (3.4).

Lemma 3.5.3. *Under Assumption 3.3.5, there exists a constant α_4 such that for all n and i,*

$$\mathrm{E} \sup_{\theta \in B(\beta,d)} |h_{ni}(s,t;\theta) - h_{ni}(s,t;\beta)| \leq \alpha_4 d^2.$$

Proof. We first note that

$$|h_{ni}(s,t;\theta) - h_{ni}(s,t;\beta)| \leq |l_{ni}(s,t;\beta) - l_{ni}(s,t;\theta)| + |Y_i(\theta) - Y_i(\beta)|.$$

From (3.4),

$$Y_i(\theta) = 1\left(Y_{i1} \leq F_1^{(-1)}\left(H_1(s;\theta_1);\theta_1\right), Y_{i2} \leq F_2^{(-1)}\left(H_2(t;\theta_2);\theta_2\right)\right).$$

As in Figure 2.2, when θ approaches β with distance d small enough, there exists integer k such that

$$M_1^k \leq X_1'\beta_1 < M_1^{k+1}$$
$$M_1^k \leq X_1'\theta_1 < M_1^k,$$

with probability 1. Hence, almost surely,

$$F_1^{(-1)}\left(H_1(s;\beta_1);\beta_1\right) = F_1^{(-1)}\left(H_1(s;\theta_1);\theta_1\right) = k.$$

By a similar argument, with small d, $F_2^{(-1)}\left(H_2(t;\beta_2);\beta_2\right) = F_2^{(-1)}\left(H_2(t;\theta_2);\theta_2\right)$. Hence, $Y_i(\theta) = Y_i(\beta)$ almost surely when d is small enough. Thus $|h_{ni}(s,t;\theta) - h_{ni}(s,t;\beta)| \leq |l_{ni}(s,t;\beta) - l_{ni}(s,t;\theta)|$. Recall (3.38) in the proof of Lemma 3.5.2, one obtains when d is

small enough, there exists constant α_4 such that for all n, i and (s,t),

$$\mathrm{E} \sup_{\theta \in B(\beta,d)} |h_{ni}(s,t;\theta) - h_{ni}(s,t;\beta)| \le \alpha_4 d^2,$$

as required. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Lemma 3.5.4. *With Assumptions 3.3.5 and 3.3.6,*

$$\frac{1}{\sqrt{n\epsilon_n^2}} \sum_{i=1}^{n} \left[h_{ni}(s,t;\hat{\beta}) - h_{ni}(s,t;\beta) \right] \to_p 0 \tag{3.39}$$

Proof. Denote

$$R_{ni}(\theta;s,t) = h_{ni}(s,t;\theta) - h_{ni}(s,t;\beta)$$

$$= 1\left(H_i(s,t;\theta) \in A_n(s,t)\right) Y_i(\theta) - 1\left(H_i(s,t) \in A_n(s,t)\right) Y_i(\beta),$$

and

$$Q_n(\theta;s,t) = \frac{1}{\sqrt{n\epsilon_n^2}} \sum_{i=1}^{n} R_{ni}(\theta;s,t).$$

For any $\epsilon > 0$, for any $\xi > 0$, let γ_ξ be the constant in Assumption 3.3.6. To show (3.39), we check the probability

$$P\left(\left|Q_n(\hat{\beta};s,t)\right| > \epsilon\right) = P\left(\left|Q_n(\hat{\beta};s,t)\right| > \epsilon, \hat{\beta} \in B(\beta, n^{-1/2}\gamma_\xi)\right)$$

$$+ P\left(\left|Q_n(\hat{\beta};s,t)\right| > \epsilon, \hat{\beta} \notin B(\beta, n^{-1/2}\gamma_\xi)\right)$$

$$\le P\left(\sup_{\theta \in B(\beta, n^{-1/2}\gamma_\xi)} |Q_n(\theta;s,t)| > \epsilon\right) + P\left(\hat{\beta} \notin B(\beta, n^{-1/2}\gamma_\xi)\right)$$

$$:= I_1 + I_2$$

By Assumption 3.3.6, $I_2 < \xi/2$.

Now define $L_{ni}(s,t) = \sup_{\theta \in B(\beta, n^{-1/2}\gamma_\xi)} |R_{ni}(\theta; s, t)|$. By Lemma 3.5.3,

$$\mathrm{E}L_{ni}(s,t) \leq \mathrm{E}\left(\sup_{\theta \in B(\beta, n^{-1/2}\gamma_\xi)} |h_{ni}(s,t;\theta) - h_{ni}(s,t;\theta)|\right)$$
$$\leq \alpha_4(\gamma_\xi/\sqrt{n})^2. \tag{3.40}$$

At the same time,

$$\sup_{\theta \in B(\beta, n^{-1/2}\gamma_\xi)} |Q_n(\theta; s, t)| \leq \frac{1}{\sqrt{n\epsilon_n^2}} \sum_{i=1}^{n} L_{ni}(s,t),$$

hence we have

$$P\left(\sup_{\theta \in B(\beta, n^{-1/2}\gamma_\xi)} |Q_n(\theta; s, t)| > \epsilon\right)$$
$$\leq P\left(\frac{1}{\sqrt{n\epsilon_n^2}} \sum_{i=1}^{n} L_{ni}(s,t) > \epsilon\right)$$
$$\leq P\left(\frac{1}{\sqrt{n\epsilon_n^2}} \sum_{i=1}^{n} [L_{ni}(s,t) - \mathrm{E}L_{ni}(s,t)] + \frac{1}{\sqrt{n\epsilon_n^2}} \sum_{i=1}^{n} \mathrm{E}L_{ni}(s,t) > \epsilon\right).$$

It follows (3.40) that
$$\frac{1}{\sqrt{n\epsilon_n^2}} \sum_{i=1}^{n} \mathrm{E}L_{ni}(s,t) \leq \alpha_4 \gamma_\xi^2 \frac{1}{\sqrt{n\epsilon_n^2}}.$$

Therefore, when n is large enough,

$$\frac{1}{\sqrt{n\epsilon_n^2}} \sum_{i=1}^{n} \mathrm{E}L_{ni}(s,t) < \epsilon/2.$$

Noting that

$$\mathrm{Var}\left(\frac{1}{\sqrt{n\epsilon_n^2}} \sum_{i=1}^{n} [L_{ni}(s,t) - \mathrm{E}L_{ni}(s,t)]\right) = \frac{1}{n\epsilon_n^2} \sum_{i=1}^{n} \left(\mathrm{E}L_{ni}(s,t)^2 - [\mathrm{E}L_{ni}(s,t)]^2\right).$$

Since $|R_{ni}(\theta; s, t)| \leq 1$, one obtains $L_{ni}(s, t)^2 \leq L_{ni}(s, t)$. From (3.40),

$$\text{Var}\left(\frac{1}{\sqrt{n\epsilon_n^2}}\sum_{i=1}^n [L_{ni}(s, t) - \mathrm{E}L_{ni}(s, t)]\right) \leq \frac{1}{n\epsilon_n^2}\alpha_4\gamma_\xi^2.$$

By Chebyshev's inequality

$$P\left(\frac{1}{\sqrt{n\epsilon_n^2}}\sum_{i=1}^n [L_{ni}(s, t) - \mathrm{E}L_{ni}(s, t)] > \epsilon/2\right)$$

$$\leq \frac{1}{(\epsilon/2)^2}\text{Var}\left(\frac{1}{\sqrt{n\epsilon_n^2}}\sum_{i=1}^n [L_{ni}(s, t) - \mathrm{E}L_{ni}(s, t)]\right) \to 0.$$

The theorem now follows $Q_n(\hat{\beta}; s, t) \to_p 0$. $\qquad\square$

Proof of Theorem 3.3.3

Proof. Recall from Lemma 3.5.2, we have $(n\pi\epsilon_n^2)^{-1}\sum_{i=1}^n l_{ni}(s, t; \hat{\beta})$ is also a consistent estimator of $f_{H(s,t)}(s, t)$. Following the proof of Theorem 3.3.2, replacing $\hat{f}_{H(s,t)}$ by its approximation $(n\pi\epsilon_n^2)^{-1}\sum_{i=1}^n l_{ni}(s, t; \hat{\beta})$ in (3.28) and (3.34), from Slutsky's theorem, we have

$$\sqrt{n\epsilon_n^2}\frac{\hat{m}_2(s, t)}{(n\pi\epsilon_n^2)^{-1}\sum_{i=1}^n l_{ni}(s, t; \hat{\beta})} \to_d N\left(0, \frac{R_2(K)}{f_{H(s,t)}(s, t)}C(s, t)\left[1 - C(s, t)\right]\right), \qquad (3.41)$$

and

$$\sqrt{n\epsilon_n^2}\left(\frac{\hat{m}_1(s, t)}{(n\pi\epsilon_n^2)^{-1}\sum_{i=1}^n l_{ni}(s, t; \hat{\beta})} - \kappa_2\zeta(s, t)\epsilon_n^2\frac{f_{H(s,t)}(s, t)}{(n\pi\epsilon_n^2)^{-1}\sum_{i=1}^n l_{ni}(s, t; \hat{\beta})}\right) \to_p 0.$$

Using Lemma 3.5.2 together with the fact that $n\epsilon_n^6 = O_p(1)$, we have

$$\sqrt{n\epsilon_n^2}\left(\kappa_2\zeta(s, t)\epsilon_n^2\frac{f_{H(s,t)}(s, t)}{(n\pi\epsilon_n^2)^{-1}\sum_{i=1}^n l_{ni}(s, t; \hat{\beta})} - \kappa_2\zeta(s, t)\epsilon_n^2\right) \to_p 0.$$

This leads to

$$\sqrt{n\epsilon_n^2}\left(\frac{\hat{m}_1(s, t)}{(n\pi\epsilon_n^2)^{-1}\sum_{i=1}^n l_{ni}(s, t; \hat{\beta})} - \kappa_2\zeta(s, t)\epsilon_n^2\right) \to_p 0. \qquad (3.42)$$

Summing up (3.41) and (3.42), we have

$$\sqrt{n\epsilon_n^2} \left(\frac{\sum_{i=1}^{n} h_{ni}(s,t;\beta)}{\sum_{i=1}^{n} l_{ni}(s,t;\hat{\beta})} - C(s,t) - \kappa_2\zeta(s,t)\epsilon_n^2 \right) \to_d N\left(0, \frac{C(s,t)\left[1 - C(s,t)\right]R_2(K)}{f_{H(s,t)}(s,t)} \right).$$

(3.43)

It follows Lemmas 3.5.2 and 3.5.4 that

$$\sqrt{n\epsilon_n^2} \left(\frac{\sum_{i=1}^{n} h_{ni}(s,t;\hat{\beta})}{\sum_{i=1}^{n} l_{ni}(s,t;\hat{\beta})} - \frac{\sum_{i=1}^{n} h_{ni}(s,t;\beta)}{\sum_{i=1}^{n} l_{ni}(s,t;\hat{\beta})} \right) \to_p 0.$$

(3.44)

Summing up (3.43) and (3.44) yields

$$\sqrt{n\epsilon_n^2} \left(\hat{C}(s,t;\hat{\beta}) - C(s,t) - \kappa_2\zeta(s,t)\epsilon_n^2 \right) \to_d N\left(0, \frac{C(s,t)\left[1 - C(s,t)\right]R_2(K)}{f_{H(s,t)}(s,t)} \right),$$

as required. $\square$

Chapter 4

Simulation

In this section, we evaluate the overall ability of the estimator to identify a copula under different scenarios. In addition, we investigate two practical issues using simulated data. The first experiment explores our nonparametric estimator working as a diagnostic tool for choosing a parametric copula. The second experiment concerns bandwidth selection.

4.1 Simulation Study Design

The simulations are conducted under scenarios with combinations of different levels of dependence and discreteness in margins. We demonstrate explicit results for Poisson outcomes. Although not reported here, we conducted the simulation with other marginal models, and the conclusions are consistent.

Marginally, for $j = 1, 2$, the mean is based on the function $E(Y_j|X_j) = \lambda_j = \exp(\beta_{j0} + X_j\beta_{j1})$, where $X_1 \sim N(0, 1)$, $X_2 \sim N(0, 1)$, independently. As indicated in Nikoloulopoulos (2013), it is more problematic to apply copulas when data are highly discrete with large probability of ties. Hence, we consider three marginal scenarios to explore the influence of the discreteness on the estimator. When λ_j is small, Y_j takes on a small number of values with high level of discreteness, while Y_j behaves analogously to a continuous variable with large λ_j. Parameters $\beta_{j0}, j = 1, 2$ are allowed to vary to obtain different marginal mean

levels:

- Small mean: $\beta_{10} = -2, \beta_{11} = 2, \beta_{20} = -2, \beta_{21} = 1.5$.

- Medium mean: $\beta_{10} = 0, \beta_{11} = 2, \beta_{20} = 0, \beta_{21} = 1.5$.

- Large mean: $\beta_{10} = 5, \beta_{11} = 2, \beta_{20} = 5, \beta_{21} = 1.5$.

Meanwhile, three levels of dependence are considered. To compare across different copulas, we quantify dependence using Kendall's τ as 0.07 for low dependence, 0.2 for moderate dependence, and 0.6 for high dependence, respectively. We also conducted the analysis on negative correlated data and found out it is the level instead of the sign of the correlation that influences the results mostly. We use sample sizes $n = 1000$ and $n = 5000$. The number of replications in each simulation is 500. The Epanechnikov kernel is used throughout.

4.2 Finite-Sample Performance

We first assess the finite sample performance of our estimator under different scenarios. Here we employ Gaussian copulas as the underlying dependence models. There are many possibilities for the dependence models. Although their results are not reported here, we can draw consistent conclusions. With simulated data, we first fit their marginal models and then plug the estimates in the copula estimator (3.5).

Figure 4.1 displays the estimator under different scenarios with sample size $n = 1000$. For clarification, the corresponding confidence intervals are given for every other copula value. The leftmost plots correspond to the cases with small marginal means. We can see both bias and variance are large under high discreteness level. Recall that in Figure 2.2, when the means of the margins are small, for a fixed point s, the corresponding $H_j(s; X_j, \beta_j)$ is distributed around 1. In this case, the number of effective observations with positive weights in (3.5) is small. As a result, large bias and variance are expected. It also has a

noticeable increase in bias towards the upper and right edges of the unit square, which was studied theoretically in Fan and Gijbels (1992).

As the marginal means increase to the medium level, as displayed in the middle column of Figure 4.1, it is clear that the accuracy of the estimator improves with smaller bias and variance. When the marginal means increase to the large scenario (right column of Figure 4.1), the estimator appears to perform well with negligible bias and variance. Indeed, as the grid points become dense, the estimator behaves similarly to the empirical copula estimator (3.3) used for continuous outcomes with order of convergence $\sqrt{n}$; see Figure 4.3 for comparison.

By comparing across different levels of dependence corresponding to the rows in Figure 4.1, we can conclude the level of dependence is less influential on the performance of the copula estimator than the discreteness. Figure 4.2 shows the results with sample size $n = 5000$. As anticipated, the bias and variance are smaller with larger sample size. This phenomenon suggests the identification of copulas even when outcomes are highly discrete is possible if the sample size is sufficiently large.

Figure 4.4 illustrates the performance of the estimator on a single sample under moderate dependence. We display the curves for the samples with 10th, 50th, and 90th percentile rank in terms of integrated squared error (ISE) among the 500 replications. The ISE is considered to be a good criterion when one wants to measure how good an estimator is for a given dataset and is defined as

$$\text{ISE}\left(\hat{C}(s,t;\hat{\beta})\right) = \int_{s,t}\left(\hat{C}(s,t;\hat{\beta}) - C(s,t)\right)^2 \mathrm{d}s\mathrm{d}t. \tag{4.1}$$

That is, a good estimator is supposed to have a small ISE value. Hence, the curve corresponding to 10% percentile sample is supposed to be closest to the true curve while the 90% curve is farthest apart. Consistent with Figures 4.1 and 4.2, the estimator is inaccurate with high level of discreteness. As marginal means increase, the estimator gets closer to the true curve. We omit the plots under low and high dependence scenarios since the conclusions
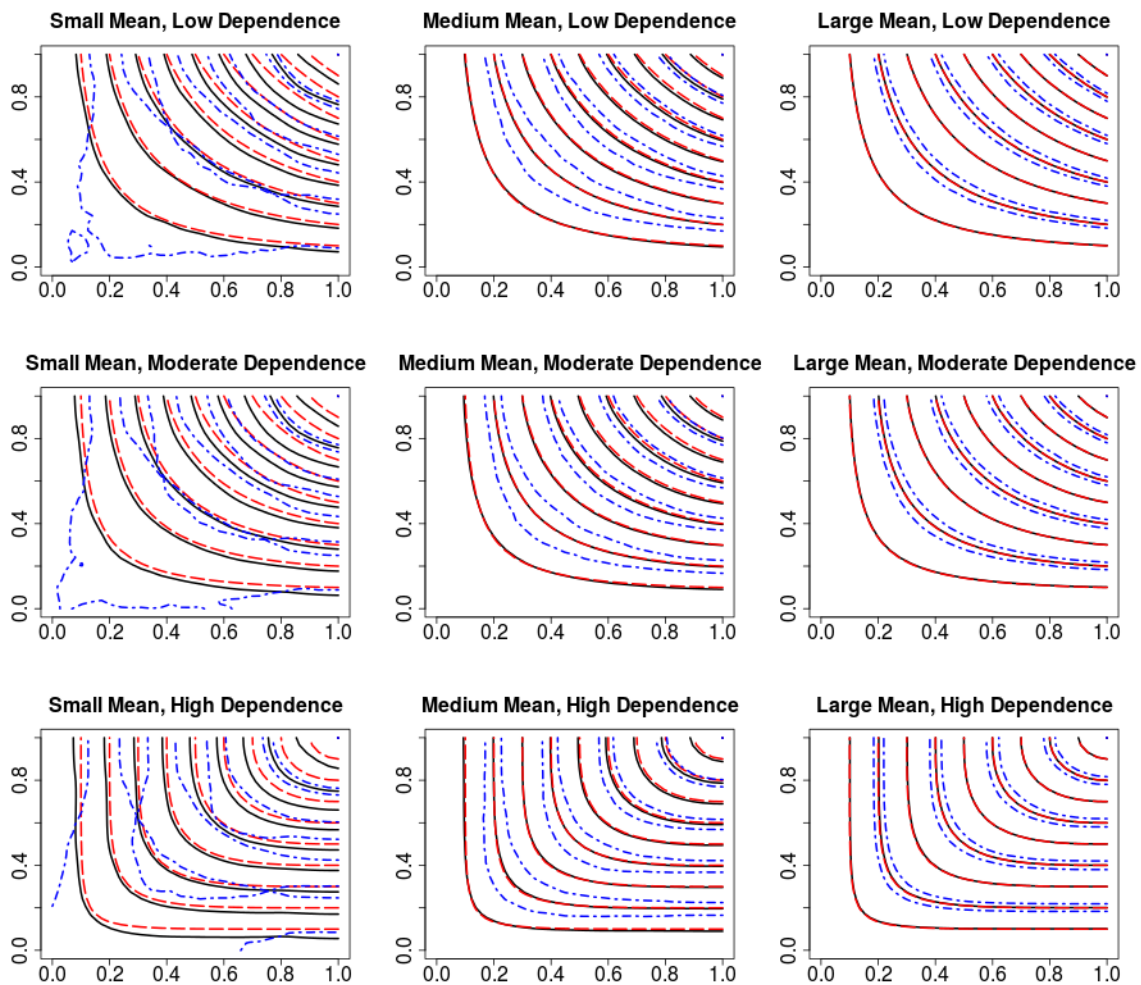
Figure 4.1: Contour plots of the nonparametric estimator under different scenarios with sample size 1000. The average of the estimator over 500 replications is given by the solid lines, while the dash-dot symbols give the corresponding 95% confidence interval for every other copula value, and the dashed lines give the underlying copulas.
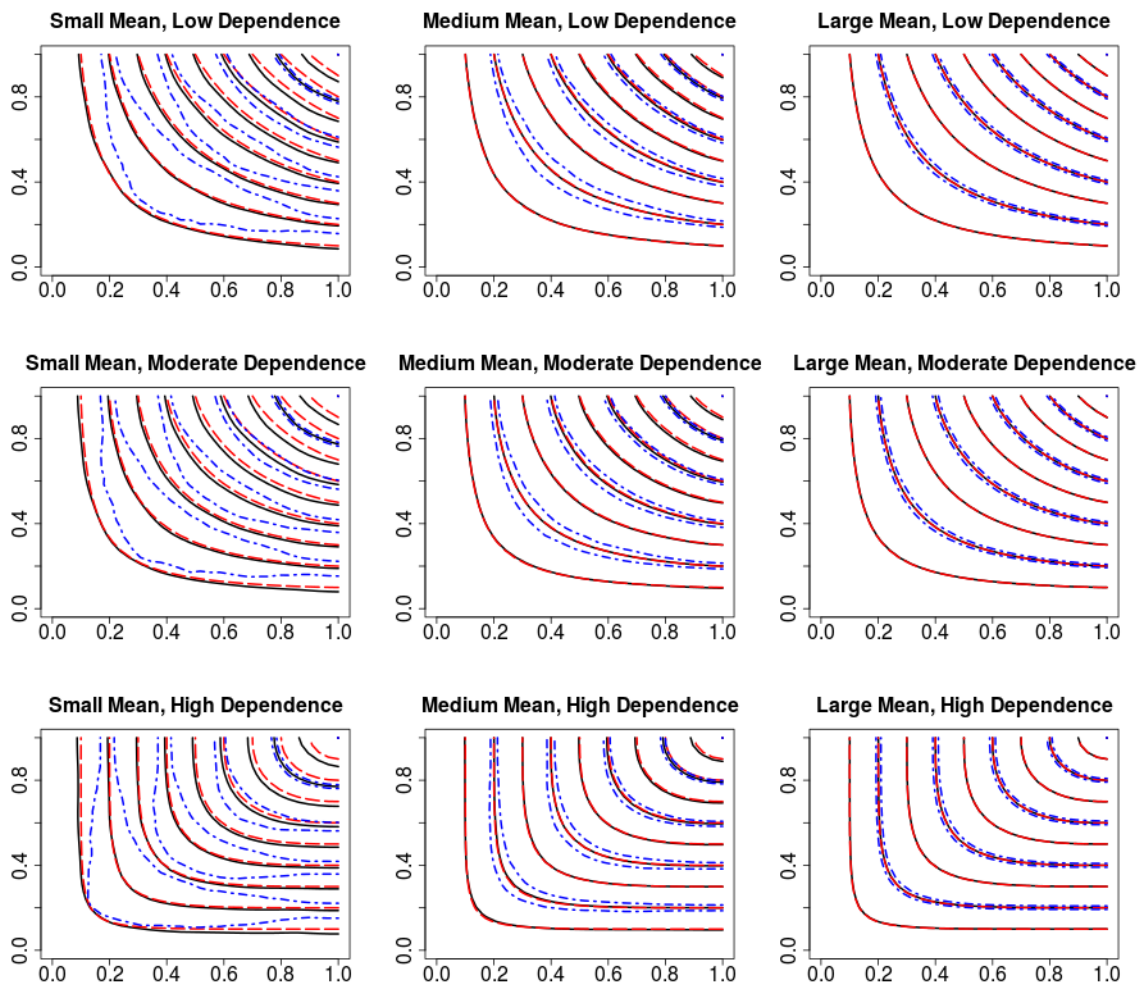
Figure 4.2: Contour plots of the nonparametric estimator under different scenarios with sample size 5000. The average of the estimator over 500 replications is given by the solid lines, while the dash-dot symbols give the corresponding 95% confidence interval for every other copula value, and the dashed lines give the underlying copulas.
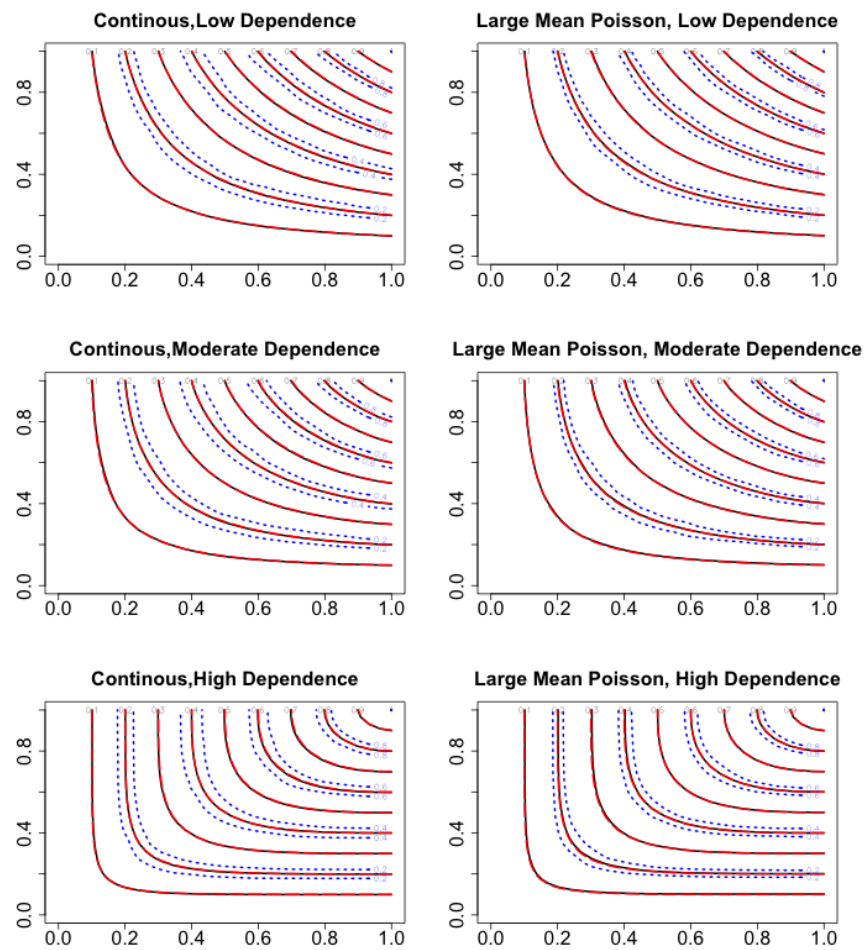
Figure 4.3: Comparison of the nonparametric estimator under large means with the empirical copula estimator for continuous outcomes. The left column displays the average of the empirical copula estimator (solid lines) with corresponding 95% confidence interval (blue dotted lines). The right column corresponds to the proposed nonparametric estimator.

are uniform.
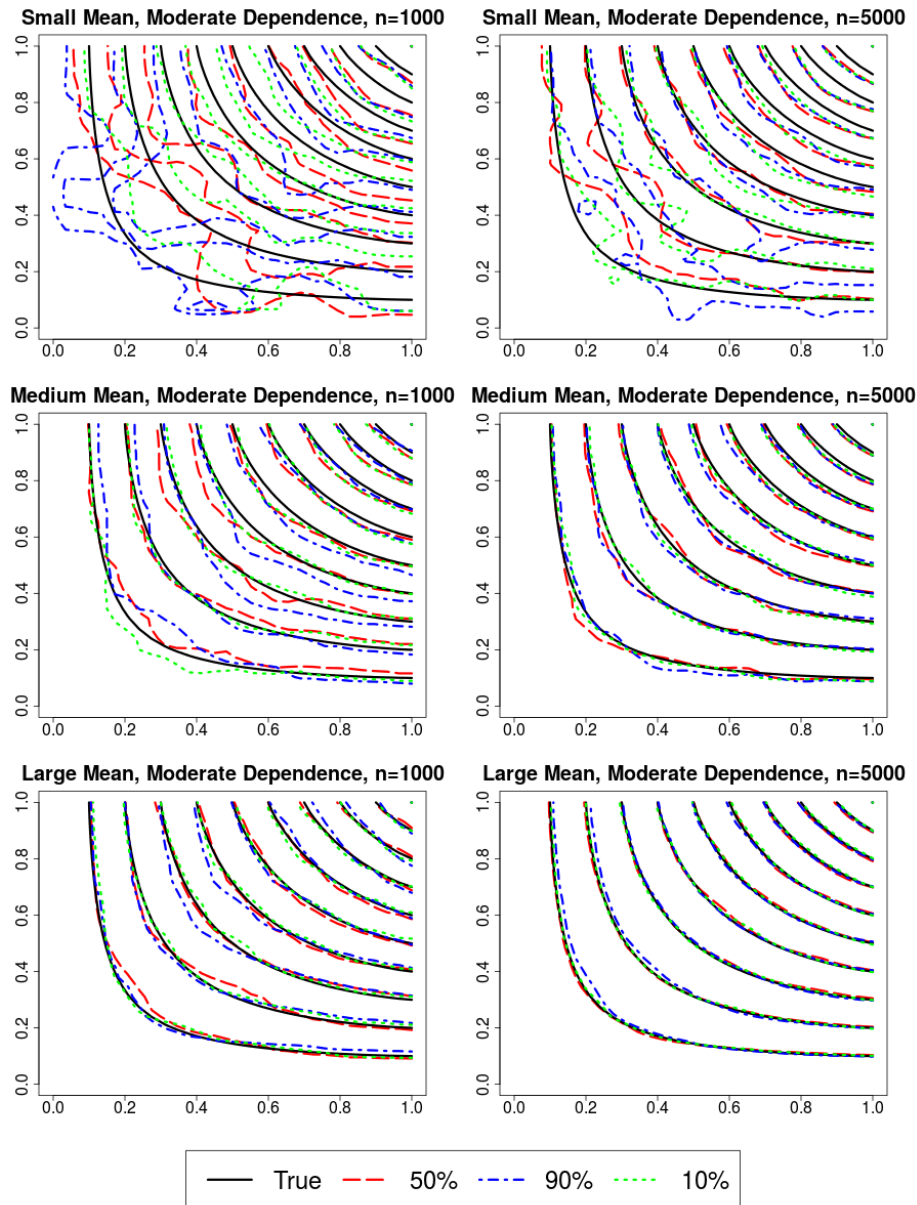
Figure 4.4: Contour plots of the nonparametric estimator for representative samples corresponding to the 10th, 50th, and 90th percentile rank in terms of the ISE

Correspondingly, the results under different scenarios are summarized numerically in Table 4.1. We quantify the performance of the estimator using the ISE defined in (4.1) and use numerical integration to compute the ISE values. As an example, when the sample size

is $n = 1000$, over the 500 replications, the average ISE of the nonparametric estimator is 2.865×10^{-3} with a standard deviation 2.081×10^{-3}. Consistent with Figure 4.1, the level of discreteness plays an important role on the performance of the nonparametric estimator, which is reflected in the ISE values. The nonparametric estimator performances better as the marginal means and the sample size increase.

Table 4.1: ISE under different scenarios (multiplied by 1000)

Marginal Mean	Dependence	n=1000		n=5000	
		Average	sd	Average	sd
Small	Low	2.865	2.081	0.857	0.311
	Moderate	3.061	2.233	0.878	0.334
	High	3.547	2.652	0.018	0.009
Medium	Low	0.331	0.118	0.107	0.030
	Moderate	0.330	0.125	0.101	0.030
	High	0.352	0.150	0.103	0.031
Large	Low	0.088	0.040	0.018	0.009
	Moderate	0.088	0.040	0.018	0.009
	High	0.091	0.047	0.019	0.010

As discussed in Section 3.3, the copula can be identified using only the probability of $(0,0)$ with $\hat{C}_0(s,t;\beta)$ defined in (3.9). Figure 4.5 displays the results for copula identification using zeros under the medium marginal mean level. Compared with the middle columns in Figures 4.1 and 4.2, it is clear that the proposed nonstandard nonparametric estimator has smaller bias and variance.

4.3   Copula Specification and Diagnosis

For those who prefer to use parametric copulas, the proposed nonparametric estimator can serve as a specification and diagnostic tool for selecting a parametric copula. We now explore the usage under different scenarios. For each of the simulations, given the generated data, we first fit the marginal models. Then, we plug the marginal estimates in (3.5) to obtain our nonparametric estimator. Meanwhile, different parametric copulas are fit through MLE. Finally, we compare the parametric copulas with our nonparametric estimator.
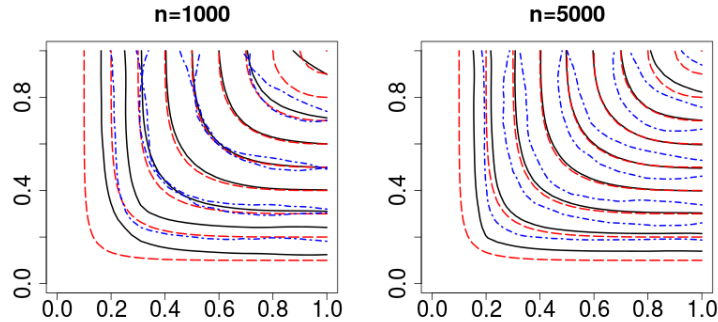
Figure 4.5: Contour plots of the nonparametric estimator using the probability of $(0,0)$. The average of the estimator over 500 replications is given by the solid lines, while the dash-dot symbols give the corresponding 95% confidence interval, and the dashed lines give the underlying copulas.

To measure the distance between the fitted parametric copulas with the nonparametric copula estimator, we use the L_2-norm distance

$$d(\hat{C}(\cdot;\hat{\beta}), \tilde{C}_{\hat{\theta}}) = \int_{s,t} (\hat{C}(s,t;\hat{\beta}) - \tilde{C}_{\hat{\theta}}(s,t))^2 \mathrm{d}s\mathrm{d}t, \tag{4.2}$$

where $\hat{C}(\cdot;\hat{\beta})$ is the proposed nonparametric estimator, and $\tilde{C}_{\hat{\theta}}$ is the parametric copula. The parametric copulas with good fitting are supposed to be close to our nonparametric estimator with small distances. Below, we generate the data using Gaussian (no tail dependencies), Clayton (lower tail dependence), and Joe (upper tail dependence) copulas to explore the impact of tail dependence.

Case 1. Gaussian copulas. We first analyze the generated data from Gaussian copulas, the most commonly used copulas without tail dependence. We include a representative graphical summary of the results under medium means in Figures 4.6 and 4.7. Due to space limitations, the results of other scenarios are summarized numerically in Table 4.2. Under low dependence, the dashed lines (corresponding to the fitted parametric copulas) across plots in Figure 4.6 are hardly distinguishable due to the fact that they are all similar to the independence copula. As a result, their distances with the nonparametric estimator

are comparable. Therefore, the choice of parametric copulas is not essential when the dependence is very weak.

In contrast, under high dependence as in Figure 4.7, we can exclude the Gumbel and Joe copulas due to the large discrepancy with the nonparametric estimator in the center of the graphs, and the Clayton copula is wide apart towards right upper corner. Recall that Gaussian and Frank copulas do not have tail dependence. Gumbel and Joe copulas have upper tail dependence, while a Clayton copula has lower tail dependence. Hence, when the dependence is strong, we can rule out copulas with wrong types of tail dependencies, and the graphical comparison with our nonparametric estimator suggests improvement. Due to the similarity in the Gaussian and Frank copulas, the choice between these two copulas is difficult and probably not that important. It is also noticeable that among copulas with upper tail dependence, the Joe copula has more significant distance than the Gumbel copula with the nonparametric estimator, which can be explained by the stronger tail dependence of the Joe copula.

Figure 4.8 displays the graphical results under small marginal means and high dependence. Due to large bias and variance in the nonparametric estimator, as demonstrated in Section 4.2, all the copulas are inside the confidence intervals. Hence, the wrong models cannot be rejected statistically, and it is hard to make conclusions about copula specification. This is consistent with Section 5.7 in Joe (2014) in which Kullback-Leibler (K-L) divergence is related to the closeness in the strength of tail dependence for positive correlated variables. They concluded that K-L divergence gets smaller with more discretization since different asymptotic rates in the joint tails do not affect rectangle probabilities in (1.4), thus a larger K-L sample size is needed to distinguish two densities.

Table 4.2 summarizes the results numerically. As an example, when the sample size is $n = 1000$, under small marginal means and low dependence, the average distance of the fitted Gaussian copula with our nonparametric estimator is 2.865×10^{-3} with standard deviation 2.081×10^{-3} over the 500 replications, while the fitted Frank copula has average distance 2.834×10^{-3} with deviance 2.095×10^{-3}. Consistent with Figure 4.8, with high level
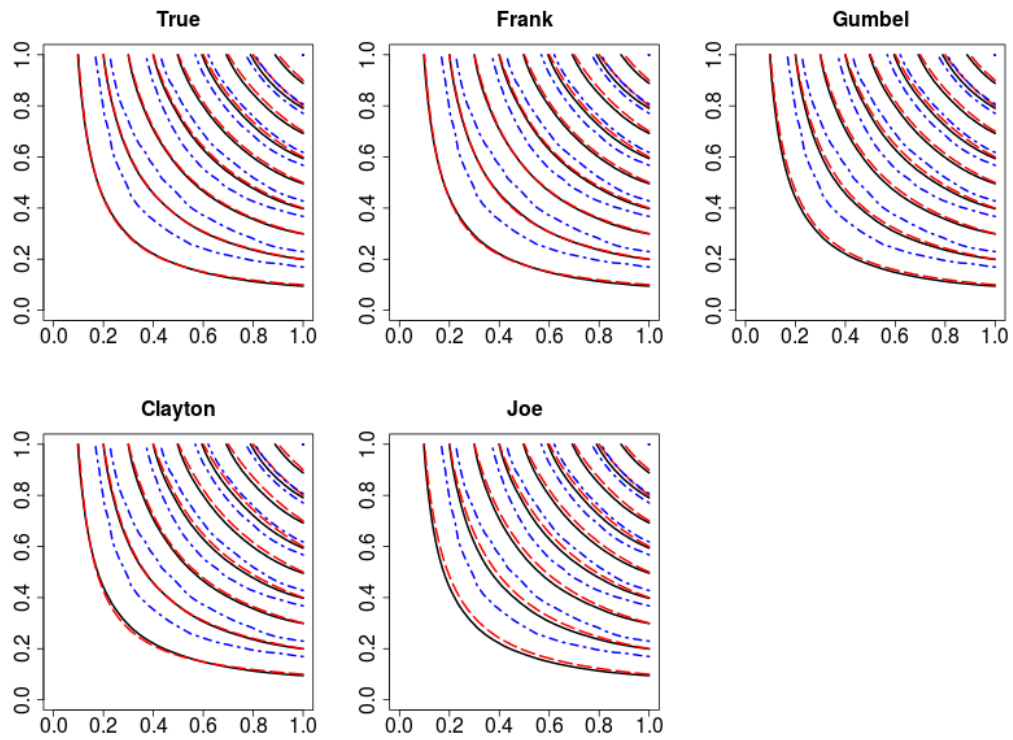
Figure 4.6: Contour plots of the nonparametric estimator compared with several parametric copulas under medium means and low dependence. The estimator is given by the solid lines, and the dash-dot symbols give the corresponding confidence intervals. The fitted parametric copulas are given by a dashed line. These plots are based on a sample size of 1000.
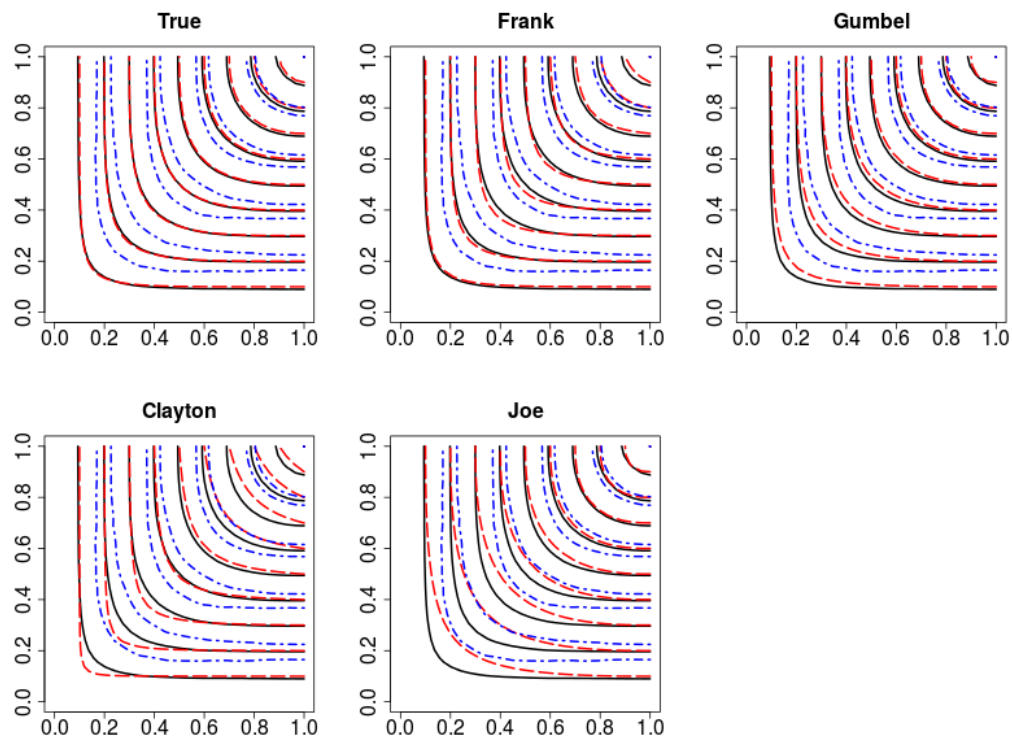
Figure 4.7: Contour plots of the nonparametric estimator compared with several parametric copulas under medium means and high dependence. The estimator is given by the solid lines, and the dash-dot symbols give the corresponding confidence intervals. The fitted parametric copulas are given by a dashed line. These plots are based on a sample size of 1000.
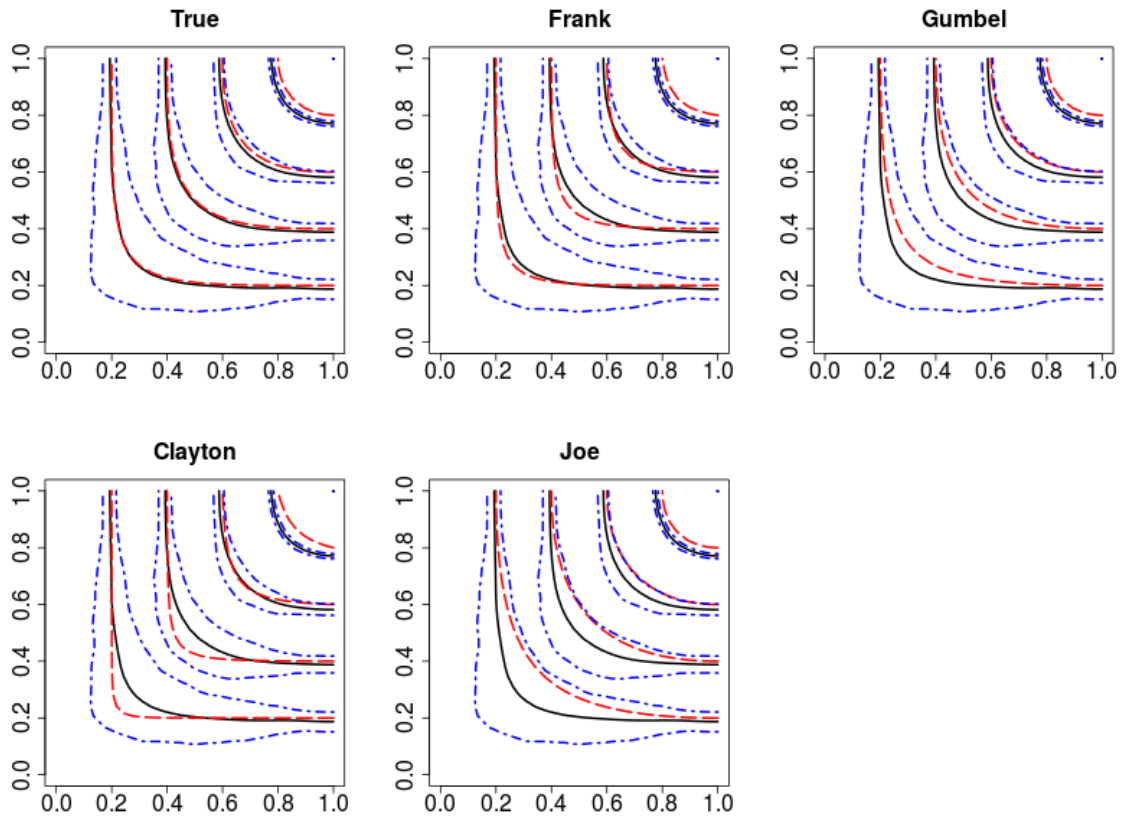
Figure 4.8: Contour plots of the nonparametric estimator compared with several parametric copulas under small means and high dependence. The estimator is given by the solid lines, and the dash-dot symbols give the corresponding confidence intervals. The fitted parametric copulas are given by a dashed line. These plots are based on a sample size of 5000.

of discreteness, the distances between different parametric copulas with the nonparametric estimator are high and comparable (first three rows of each block). Thus, it is difficult to pick up a copula. When the marginal means are at medium and large levels, the strength of dependence plays an important role in the model specification. Under low dependence, we are unable to distinguish most of the copulas in terms of the distance, except the Joe copula shows worse fitting at large mean level. With stronger dependence, we can rule out the Gumbel, Joe, and Clayton copulas, especially under high dependence, where the true model outperforms alternative models clearly. Again, the Gaussian and Frank copulas are generally indistinguishable, except that the difference is more noticeable with large marginal means and high dependence.

Case 2. Clayton copulas. To further explore the impact of tail dependence, we next consider copulas with lower tail dependence. Table 4.3 portrays the results of the Clayton copulas. As we concluded the choice of copulas is not essential under low dependence or small marginal means, we omit the corresponding results here. In all the scenarios, the true model has smallest distance with the nonparametric estimator. Meanwhile, we can rule out the Gumbel and Joe copulas easily as they have opposite tail dependence structures of the Clayton copulas. The Frank copulas are far apart when the dependence is high.

Case 3. Copulas with upper tail dependence. Now we use copulas with upper tail dependence as underlying models. Here we employ two copulas to check if the copula with same type of tail dependence always outperforms other copulas.

Table 4.4 exhibits surprising yet practical results when the data are generated using Gumbel copulas. The true model has smallest distance. However, the Frank copulas are closer to our nonparametric estimator than the Joe copulas, which is counterintuitive as Joe and Gumbel copulas both have upper tail dependence. To assist with the interpretation of the results, Figure 4.9 displays the contour plots of the fitted parametric copulas for a dataset generated by a Gumbel copula with large means and high dependence. The fitted

Table 4.2: Distances of different parametric copulas with the nonparametric estimator when Gaussian copulas are the underlying (multiplied by 1000)

n	Marginal Mean	Dependence	True		Frank		Gumbel		Clayton		Joe	
			Average	sd	Average	sd	Average	sd	Average	sd	Average	sd
1000	Small	Low	2.865	2.081	2.834	2.095	3.026	2.266	2.809	1.928	3.092	2.309
		Moderate	3.061	2.233	2.985	2.151	3.523	2.635	3.045	1.853	3.855	2.853
		High	3.547	2.652	3.495	2.519	4.049	3.205	3.633	2.288	4.994	3.893
	Medium	Low	0.331	0.118	0.323	0.102	0.356	0.139	0.351	0.114	0.394	0.169
		Moderate	0.330	0.125	0.330	0.110	0.409	0.187	0.433	0.146	0.587	0.282
		High	0.352	0.150	0.392	0.158	0.432	0.221	0.552	0.190	0.846	0.408
	Large	Low	0.088	0.040	0.081	0.033	0.090	0.039	0.096	0.049	0.115	0.055
		Moderate	0.088	0.040	0.087	0.035	0.112	0.051	0.174	0.085	0.236	0.098
		High	0.091	0.047	0.114	0.052	0.128	0.064	0.425	0.144	0.459	0.149
5000	Small	Low	0.857	0.311	0.843	0.291	0.953	0.391	0.863	0.261	1.000	0.423
		Moderate	0.878	0.334	0.849	0.285	1.171	0.537	0.986	0.281	1.425	0.661
		High	0.974	0.429	1.014	0.402	1.286	0.682	1.204	0.445	2.039	1.054
	Medium	Low	0.107	0.030	0.104	0.025	0.126	0.046	0.124	0.031	0.157	0.064
		Moderate	0.101	0.030	0.107	0.025	0.165	0.069	0.189	0.043	0.325	0.118
		High	0.103	0.031	0.142	0.040	0.171	0.078	0.281	0.060	0.565	0.177
	Large	Low	0.018	0.009	0.017	0.007	0.026	0.012	0.032	0.016	0.050	0.023
		Moderate	0.018	0.009	0.023	0.009	0.048	0.020	0.110	0.035	0.173	0.046
		High	0.019	0.010	0.043	0.016	0.059	0.023	0.358	0.061	0.393	0.069

Table 4.3: Distances of different parametric copulas with the nonparametric estimator when Clayton copulas are the underlying (multiplied by 1000)

n	Marginal Mean	Dependence	True		Frank		Gumbel		Joe	
			Mean	sd	Mean	sd	Mean	sd	Mean	sd
1000	Medium	Moderate	0.332	0.121	0.364	0.159	0.629	0.292	0.929	0.392
		High	0.349	0.146	0.430	0.236	1.131	0.519	2.287	0.792
	Large	Moderate	0.089	0.040	0.134	0.055	0.224	0.089	0.487	0.161
		High	0.091	0.048	0.202	0.081	0.483	0.165	1.511	0.340
5000	Medium	Moderate	0.106	0.030	0.143	0.049	0.353	0.120	0.613	0.173
		High	0.113	0.035	0.204	0.073	0.859	0.224	1.974	0.361
	Large	Moderate	0.018	0.008	0.075	0.024	0.164	0.037	0.423	0.071
		High	0.019	0.010	0.139	0.034	0.420	0.073	1.448	0.156

Joe copula has bigger tail dependence coefficient 0.75 than 0.69 of the Gumbel copula. Thus, the curve of the Gumbel copula is between the Joe and the Frank copula in Figure 4.9. Moreover, the curve of the Gumbel copula is closer to the Frank copula than the Joe copula. As a result, the Frank copula is preferable to the Joe copula.
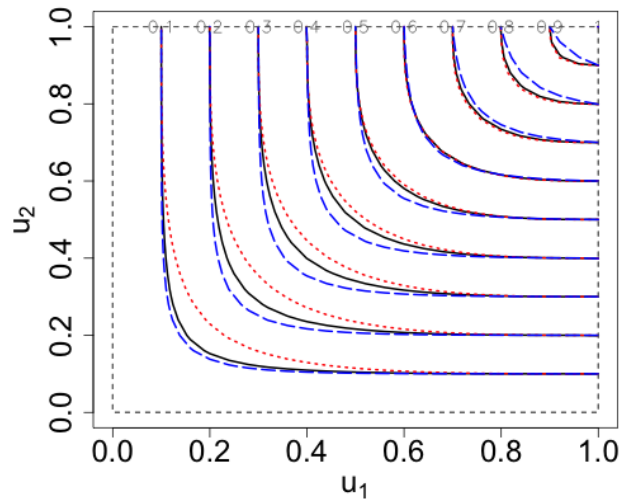
Figure 4.9: Contour plots of fitted parametric copulas. The dataset was generated by a Gumbel copula with large means and high dependence. The black solid lines give the underlying Gumbel copula while the fitted Frank copula is given by blue dashed lines, and the fitted Joe copula is given by red dotted lines.

For the other case, Table 4.5 shows the results when the Joe copula is the data generating mechanism. As explained earlier, the Gumbel copula is between the Joe and Frank copulas. Hence the Gumbel copula has smaller distance than the Frank copula.

To summarize, first, the selection of copula is more important with large marginal means and high dependence. Second, overall, our nonparametric estimator is likely to exclude copulas with wrong tail behaviors, especially those with opposite tail dependence structures of the underlying model. However, sometimes copulas without tail dependence might be preferred to the copulas with correct tail dependence type. In the situations where it seems ambiguous between copulas, we suggest expanding the candidate pool.

Table 4.4: Distances of different parametric copulas with the nonparametric estimator when Gumbel copulas are the underlying (multiplied by 1000)

n	Marginal Mean	Dependence	True Mean	sd	Frank Mean	sd	Clayton Mean	sd	Joe Mean	sd
1000	Medium	Moderate	0.320	0.123	0.363	0.114	0.528	0.150	0.380	0.169
		High	0.334	0.144	0.421	0.144	0.682	0.191	0.458	0.245
	Large	Moderate	0.088	0.042	0.112	0.048	0.279	0.114	0.117	0.061
		High	0.092	0.048	0.141	0.066	0.758	0.191	0.184	0.084
5000	Medium	Moderate	0.096	0.028	0.157	0.046	0.277	0.056	0.141	0.057
		High	0.099	0.031	0.196	0.067	0.412	0.078	0.207	0.077
	Large	Moderate	0.019	0.009	0.047	0.017	0.209	0.051	0.050	0.020
		High	0.018	0.009	0.069	0.022	0.672	0.082	0.110	0.030

Table 4.5: Distances of different parametric copulas with the nonparametric estimator when Joe copulas are the underlying (multiplied by 1000)

n	Marginal Mean	Dependence	True Mean	sd	Frank Mean	sd	Gumbel Mean	sd	Clayton Mean	sd
1000	Medium	Moderate	0.311	0.117	0.454	0.145	0.328	0.107	0.646	0.160
		High	0.320	0.129	0.540	0.210	0.391	0.156	0.956	0.247
	Large	Moderate	0.093	0.044	0.176	0.066	0.112	0.050	0.420	0.143
		High	0.085	0.041	0.233	0.087	0.142	0.060	1.473	0.246
5000	Medium	Moderate	0.091	0.026	0.251	0.061	0.126	0.036	0.398	0.066
		High	0.092	0.027	0.318	0.087	0.174	0.054	0.694	0.098
	Large	Moderate	0.018	0.009	0.109	0.026	0.044	0.016	0.353	0.061
		High	0.019	0.009	0.164	0.037	0.075	0.022	1.414	0.125

4.4   Selection of Bandwidth

An important choice to be made is the bandwidth, which provides a trade-off between bias and variance of the estimator. In this section, we establish a data-driven selector for the bandwidth. In this section, we establish a data-driven selector for the bandwidth.

To demonstrate sensitivity of the proposed estimator to different bandwidths under different scenarios, Figure 4.10 portrays the contour plots of the nonparametric estimator with different bandwidths under different marginal mean levels. It appears that the bandwidth plays a more important role in the small and medium marginal mean settings than in the large mean cases where the estimator is not as sensitive to the selection of bandwidth. Therefore, we do not emphasize bandwidth selection for large mean cases in this section.
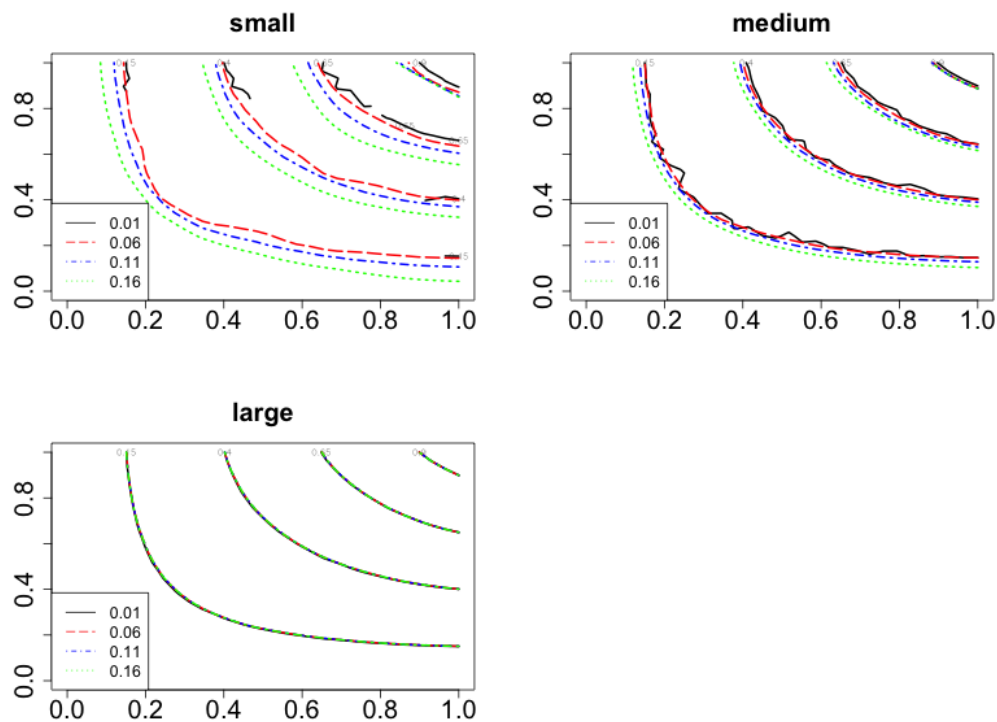
Figure 4.10: Contour plots of the nonparametric estimator with different bandwidths at moderate dependence. Sample size: 1000.

The benchmark bandwidth here is the minimizer of the ISE as defined in (4.1). However,

in practice, $C(s,t)$ is unknown. In this section, we propose a practical "plug-in" bandwidth selection rule. The independence copula is a natural choice to plug in (4.1). Furthermore, motivated by the idea of working covariance in generalized estimating equations (GEEs) (Zeger and Liang 1986), we propose a procedure to replace $C(s,t)$ by a "rule-of-thumb" estimator: a Frank copula estimated by maximum likelihood. We apply a Frank copula as the working copula for the following practical reasons. First, a Frank copula can capture a wide range of dependence including positive and negative dependence. Second, Frank copulas belong to the Archimedean family with a closed form of distribution functions and benefits of easy computation. There is no absolute justification for this choice. If there is prior information such as tail dependence, a more informative copula can be applied in our procedure. We show numerically the proposed selector is satisfactory even with misspecification in tail behavior.

The idea of "plug-in" has been widely applied for choosing smoothing parameters in kernel density estimation (Chiu 1991) and nonparametric regression with an odd number order of local polynomial (Ruppert et al. 1995), in which optimal smoothing parameters are selected by minimizing an approximation to the mean integrated squared error or its asymptotic form. Nonetheless, the asymptotic mean integrated mean squared error for the proposed nonparametric estimator involves $f_{H(s,t;\mathbf{X})}(s,t)$ and its derivatives, as in Theorem 3.3.3. Since $f_{H(s,t;\mathbf{X})}(s,t)$ is not a typical density function, its approximation takes extra efforts than plugging in a parametric density function. In addition, the estimation of the derivatives of $f_{H(s,t;\mathbf{X})}(s,t)$ is challenging and has a smaller order of convergence. We also avoid using cross-validation due to its computational burden. Because $H(s,t;\mathbf{X})$ varies with (s,t), the typical setting of generalized cross-validation, which is an approximation to cross-validation and commonly applied for bandwidth selection in nonparametric regression, is violated.

We conduct a simulation study to assess the proposed bandwidth selector by comparing it with the benchmark selector minimizing the ISE values. We use numerical integration to compute the ISE. In addition, we include the results using the independence copula as

working copulas in our procedure. Tables 4.6, 4.7, and 4.8 report the numerical results. We compare the selected bandwidths from different selectors and the resulted ISE values. We do not concern the low dependence scenarios here, since the independence copula is close to the truth in these cases and can be used as the working copula without doubt.

Table 4.6 shows the result with Gaussian copulas as the underlying dependence structures. For example, when the data are generated with small marginal means and a Gaussian copula at moderate dependence level, the minimizer of the ISE value gives a bandwidth 90.653×10^{-3} on average with standard deviation 12.440×10^{-3}. With the selected bandwidths, the ISE values of our nonparametric estimator are 2.703×10^{-3} on average with standard deviation 1.202×10^{-3}. We see that the proposed procedure returns bandwidths close to the results of the benchmark selector across different marginal means, dependence levels, and sample sizes, while using the independence copula tends to undersmooth significantly, especially when there is high dependence. Intuitively, the discrepancy between the underlying copula and independence is large under high dependence. Therefore, we suggest not using the independence copula alternative for bandwidth selection when significant dependence is detected.

While Gaussian and Frank copulas have same tail dependence properties, to evaluate the proposed bandwidth selector when the underlying copula has different tail dependencies, we conduct the simulation using Gumbel and Clayton copulas to generate the data. As portrayed in Tables 4.7 and 4.8, it is not surprising that the proposed selector performs not as good as when Gaussian copulas are the underlying models (Table 4.6). For Gumbel copulas, the selector using a Frank copula gives larger bandwidths while for Clayton copulas the bandwidths are smaller than the benchmark values. However, we think the selector performs satisfactorily even with misspecification of tail dependence, which is reflected in the ISE values of the resulted copulas.

Table 4.6: Bandwidths and resulting ISE values of different selectors for data generated by Gaussian copulas (multiplied by 1000)

| Dependence | n | | Small Mean | | | | | | Medium Mean | | | | | |
| | | | Benchmark | | Frank | | Independent | | Benchmark | | Frank | | Independent | |
			Average	sd	Average	sd	Average	sd	Average	sd	Average	sd	Average	sd
Moderate	1000	h	90.653	12.440	90.989	12.871	77.347	9.392	74.132	11.369	74.500	11.429	55.968	7.728
		ISE	2.703	1.202	2.718	1.201	2.855	1.170	0.330	0.119	0.333	0.119	0.390	0.148
	5000	h	67.095	5.636	68.653	5.783	55.768	4.723	53.732	4.986	53.142	5.424	38.626	7.887
		ISE	0.852	0.335	0.856	0.335	0.937	0.331	0.096	0.026	0.096	0.026	0.127	0.044
High	1000	h	98.779	15.995	102.021	16.831	73.516	10.405	77.189	13.036	77.226	13.972	56.447	9.446
		ISE	2.904	1.392	2.915	1.391	3.531	1.463	0.329	0.123	0.330	0.123	0.414	0.168
	5000	h	73.495	7.404	76.653	8.910	52.695	7.960	56.532	5.963	55.058	6.723	38.111	14.678
		ISE	0.929	0.432	0.935	0.431	1.234	0.432	0.097	0.028	0.097	0.028	0.197	0.167

Table 4.7: Bandwidths and resulting ISE values of different selectors for data generated by Gumbel copulas (multiplied by 1000)

| Dependence | n | | Small Mean | | | | | | Medium Mean | | | | | |
| | | | Benchmark | | Frank | | Independent | | Benchmark | | Frank | | Independent | |
			Average	sd	Average	sd	Average	sd	Average	sd	Average	sd	Average	sd
Moderate	1000	h	91.453	10.994	98.232	11.429	78.947	10.112	74.979	10.713	79.032	10.988	57.332	8.801
		ISE	2.801	1.588	2.861	1.580	2.945	1.565	0.318	0.127	0.326	0.128	0.371	0.130
	5000	h	68.189	6.480	75.895	7.928	55.347	5.285	55.021	5.507	56.200	8.418	37.042	8.014
		ISE	0.812	0.392	0.847	0.388	0.915	0.385	0.092	0.025	0.094	0.025	0.135	0.052
High	1000	h	99.032	14.815	105.011	16.155	75.242	13.271	79.768	12.244	82.753	15.498	58.989	11.332
		ISE	3.313	2.172	3.349	2.162	3.888	2.116	0.334	0.122	0.338	0.122	0.414	0.149
	5000	h	73.958	8.385	80.021	10.893	52.189	8.202	58.521	6.583	58.521	10.046	37.742	14.240
		ISE	0.849	0.317	0.874	0.314	1.172	0.403	0.096	0.031	0.098	0.032	0.197	0.132

Table 4.8: Bandwidths and resulting ISE values of different selectors for data generated by Clayton copulas (multiplied by 1000)

Dependence	n		Small Mean						Medium Mean					
			True		Frank		Independent		True		Frank		Independent	
			Average	sd	Average	sd	Average	sd	Average	sd	Average	sd	Average	sd
Moderate	1000	h	93.305	11.133	89.726	10.153	81.600	9.662	74.795	11.602	72.511	9.739	58.032	8.908
		ISE	2.973	1.683	3.003	1.676	3.110	1.652	0.326	0.125	0.331	0.125	0.376	0.135
	5000	h	68.274	6.106	65.579	5.845	57.705	6.108	54.395	6.795	53.695	5.902	39.732	8.234
		ISE	0.895	0.309	0.902	0.307	0.968	0.308	0.102	0.027	0.103	0.027	0.133	0.046
High	1000	h	100.253	16.051	94.989	15.740	76.926	15.293	78.037	12.718	75.753	11.164	57.884	10.681
		ISE	3.570	3.838	3.606	3.831	4.170	3.750	0.334	0.154	0.337	0.154	0.413	0.171
	5000	h	73.326	9.430	69.411	8.260	54.042	8.277	56.826	7.825	57.637	7.050	39.768	14.752
		ISE	1.006	0.388	1.021	0.386	1.287	0.434	0.105	0.030	0.107	0.030	0.196	0.135

Chapter 5

Data Analysis

To illustrate the nonparametric estimator on real data, we use our model to investigate the dependence of insurance claim frequencies across different business lines using a unique dataset from the Local Government Property Insurance Fund (LGPIF) in the state of Wisconsin.

The LGPIF was established to provide property insurance for local government entities that include counties, cities, towns, villages, school districts, fire departments, and other miscellaneous entities. The fund provides different types of coverage including government buildings, vehicles, and equipments. For example, a county may need coverage for the buildings (and their contents) that it owns as well as coverage for its automobiles and trucks. The LGPIF operates similarly to a typical insurer, hence the data provide a good example for multi-line insurance companies encountered in practice. More details about the project may be found at the Local Government Property Insurance Fund project website `https://sites.google.com/a/wisc.edu/local-government-property-insurance-fund/`.

The data consist of six coverage groups: building and content (BC), contractor's equipment (IM), comprehensive new (PN), comprehensive old (PO), collision new (CN), and collision old (CO) coverage. The data are longitudinal from year 2006 to 2011.

5.1 Collaborative Work

Using the LGPIF data, I have done a few collaborative projects summarized as follows.

5.1.1 Multivariate Frequency-Severity Regression Models in Insurance

In insurance and related industries including healthcare, it is common to have several outcome measures that the analyst wishes to understand using explanatory variables. My collaborative paper with Professor Frees and my fellow doctoral student Gee Lee (Frees et al. 2016) synthesizes and extends the literature on multivariate frequency-severity regression modeling. This paper contributes to this body of literature by focusing on the use of a copula for modeling the dependence among these outcomes; a major advantage of this tool is that it preserves the body of work established for marginal models.

The identification and estimation procedures has three components in a copula regression model:

1. Fit the mean structure.

2. Fit the variance structure with a selected distribution.

3. Fit the dependence structure with a choice of copula.

Marginally, we fit frequency and loss amounts. Then copulas are used to model non-linear associations. Dependence is fit at two levels. The first is between frequency and average severity within each line. The second is among different lines.

Frequency. Since there are inflated numbers of 0 s and 1 s in LGPIF data, a "zero-one-inflated" model was introduced. As an extension of the zero-inflated method, a zero-one-inflated model employs two generating processes. The first process is governed by a multinomial distribution that generates structural zeros and ones. The second process is governed by a Poisson or negative binomial distribution that generates counts, some of which may be zero or one.

Denote the latent variable in the first process as $I_i, i = 1, \ldots, n$, which follows a multinomial distribution with possible values 0, 1, and 2 with corresponding probabilities $\pi_{0,i}, \pi_{1,i}, \pi_{2,i} = 1 - \pi_{0,i} - \pi_{1,i}$. Here, N_i is frequency and

$$N_i \sim \begin{cases} 0 & I_i = 0 \\ 1 & I_i = 1 \\ P_i & I_i = 2, \end{cases}$$

where P_i may be a Poisson or negative binomial distribution. With this, the probability mass function of N_i is

$$f_{N,i}(n) = \pi_{0,i} I_{\{n=0\}} + \pi_{1,i} I_{\{n=1\}} + \pi_{2,i} P_i(n).$$

A logit specification is used to parameterize the probabilities for the latent variable I_i. Denote the covariates associated with I_i as $\mathbf{z}_i$. Using level 2 as a reference, the specification is

$$log\frac{\pi_{j,i}}{\pi_{2,i}} = \mathbf{z}_i' \boldsymbol{\gamma}_j, j = 0, 1.$$

Correspondingly,

$$\pi_{j,i} = \frac{\exp(\mathbf{z}_i'\boldsymbol{\gamma}_j)}{1 + \exp(\mathbf{z}_i'\boldsymbol{\gamma}_j) + \exp(\mathbf{z}_i'\boldsymbol{\gamma}_j)}, j = 0, 1.$$

$$\pi_{2,i} = 1 - \pi_{0,i} - \pi_{1,i}$$

MLE is used to fit the parameters.

Severity. We used a distribution family known as the "generalized beta of the second kind", or *GB2*, for short, to analyze the average severity of claims. A random variable with a *GB2* distribution can be written as

$$e^\mu \left(\frac{G_1}{G_2}\right)^\sigma = C_1 e^\mu F^\sigma = e^\mu(\frac{Z}{1-Z})^\sigma,$$

where the constant $C_1 = (\alpha_1/\alpha_2)^\sigma$, G_1 and G_2 are independent gamma random variables with scale parameter 1 and shape parameters α_1 and α_2, respectively. Further, the random variable F has an F-distribution with degrees of freedom $2\alpha_1$ and $2\alpha_2$, and the random variable Z has a beta distribution with parameters α_1 and α_2. Thus, the *GB2* family has four parameters $(\alpha_1, \alpha_2, \mu, \text{and } \sigma)$, where μ is the location parameter. The *GB2* is a flexible distribution that accommodates positive or negative skewness, as well as heavy tails.

For incorporating covariates, it is straightforward to show that the regression function is of the form

$$E(y|\mathbf{x}) = C_2 \exp(\mu(\mathbf{x})) = C_2 \exp(\mathbf{x}'\beta),$$

where the constant C_2 can be calculated with other (non-location) model parameters. Under the most commonly used way of parametrization for *GB2*, where μ is associated with covariates, if $-\alpha_1 < \sigma < \alpha_2$,

$$C_2 = \frac{B(\alpha_1 + \sigma, \alpha_2 - \sigma)}{B(\alpha_1, \alpha_2)}$$

where $B(\alpha_1, \alpha_2) = \Gamma(\alpha_1)\Gamma(\alpha_2)/\Gamma(\alpha_1 + \alpha_2)$. Thus, one can interpret the regression coefficients in terms of a proportional change. That is, $\partial[\ln E(y)]/\partial x_k = \beta_k$.

Frequency Severity Dependency Models. We emphasize that the average severity may depend on frequency, even when the classical assumption that the claim frequency is independent of claim severity holds. To see how the copula approach works, let $\bar{S}$ represent average severity of claims and N denote frequency. Using a copula, we can express the likelihood as

$$f_{\bar{S},N}(s,n) = \begin{cases} f_{\bar{S},N}(s,n|N > 0)P(N > 0) & \text{for } n > 0 \\ P(N = 0) & \text{for } n = s = 0 \end{cases}$$

Denote

$$C_1(u,v) = \frac{\partial}{\partial u} C(u,v) = P(V \leq v | U = u).$$

With this,

$$
\begin{aligned}
f_{\bar{S},N}(s,n|N>0) &= \frac{\partial}{\partial s} P(\bar{S} \leq s, N \leq n | N > 0) \\
&= \frac{\partial}{\partial s} C(F_{\bar{S}}(s), F_N(n|N>0)) \\
&= f_{\bar{S}}(s) C_1(F_{\bar{S}}(s), F_N(n|N>0)).
\end{aligned}
$$

This yields the following expression for the likelihood

$$
f_{\bar{S},N}(s,n) =
\begin{cases}
f_{\bar{S}}(s) P(N>0) (C_1(F_{\bar{S}}(s), F_N(n|N>0)) \\
\qquad\qquad -C_1(F_{\bar{S}}(s), F_N(n-1|N>0))) & \text{for } s > 0, n \geq 1 \\
P(N=0) & \text{for } s = 0, N = 0.
\end{cases}
$$

For this dataset, we found significantly strong negative association between frequency and average severity for the building and contents (BC) line.

Dependence between Different Lines. The second level of dependence lies between different lines. We fit the dependence model for frequencies, severities, and aggregate loss with Tweedie margins. A Gaussian copula was applied and the composite likelihood method is used for computation. We saw strong correlations among lines.

Model Validation. After identification and estimation, we also compared a number of alternative models based on the training and on the test samples.

For the training sample, we used a statistic due to Vuong (Vuong 1989) to do the "in-sample" comparisons. For this statistic, one calculates the contribution to the logarithmic likelihood for two models, say, $l_i^{(1)}$ and $l_i^{(2)}$. One prefers Model (1) compared to Model (2) if the average difference, $\bar{D} = m^{-1} \sum_{i=1}^{m} D_i$, is positive, where $D_i = l_i^{(1)} - l_i^{(2)}$ and m is the size of the validation sample. To assess the significance of this difference, one can apply approximate normality with approximate standard errors given as $SD_D / \sqrt{m}$ where

$SD_D^2 = (m-1)^{-1} \sum_{i=1}^{m} (D_i - \bar{D})^2$. The Gaussian copula was preferred in our frequency-severity model.

Comparison among models using test data, or "out-of-sample" comparisons are also important in insurance because many of these models are used for predictive purposes such as setting rates for new customers. We considered the nonparametric Spearman correlation between model predictions and the held-out claims as well as the Gini index to measure the satisfaction of the fund manager with each score. The zero-one-inflated negative binomial model and the long-tail distribution (*GB2*) marginals bring much improvement.

5.1.2 Pair Copula Constructions for Insurance Experience Rating

Dependence could be introduced through observations over time and contribute to predictive analytics. In non-life insurance, insurers model longitudinal insurance claims and infer the predictive distribution given previous loss experience, a process known as experience rating, to adjust the premiums for insurance. Experience rating is challenging when the distribution of claims is complicated in the way that it has a point mass zero associated with no claims and a long tail distribution of severity given occurrence of claims. In collaborative work with Professor Shi (Shi and Yang 2017), we introduced a mixed vine pair copula construction framework for modeling semicontinuous longitudinal claims. The quantity of interest is the entity-level cost of claims from building and contents insurance of the LGPIF. In addition, the fund further breaks down the total cost of claims by the cause of losses, known as peril in property insurance. In this application, we examine the total cost as well as the cost by peril.

First, we used a two-component mixture model to accommodate the marginal distributions. Specifically, each outcome is assumed as being generated from a degenerate distribution at zero with certain probability and being generated from a skewed and heavy tailed distribution defined on $(0, +\infty)$ otherwise. Separating the frequency and severity allows for different sets of predictors as well as different effects of the same predictor on each component. For the claim frequency, we considered a logit specification due to the straightforward

interpretability of model parameters, while for the claim severity, we employed the *GB2* distribution to model skewed and heavy-tailed data.

After carefully examining the marginal distributions, we introduced the vine pair copula constructions for mixed data to model the temporal dependence. A vine is a graphical model for dependent random variables and was applied on continuous ((Aas et al. 2009) and discrete outcomes (Panagiotelis et al. 2012). To model the longitudinal data, we focus on a specific vine - D-vine, which is a natural choice for temporal order. An example of a D-vine on five repeated observations is exhibited in Figure 5.1.
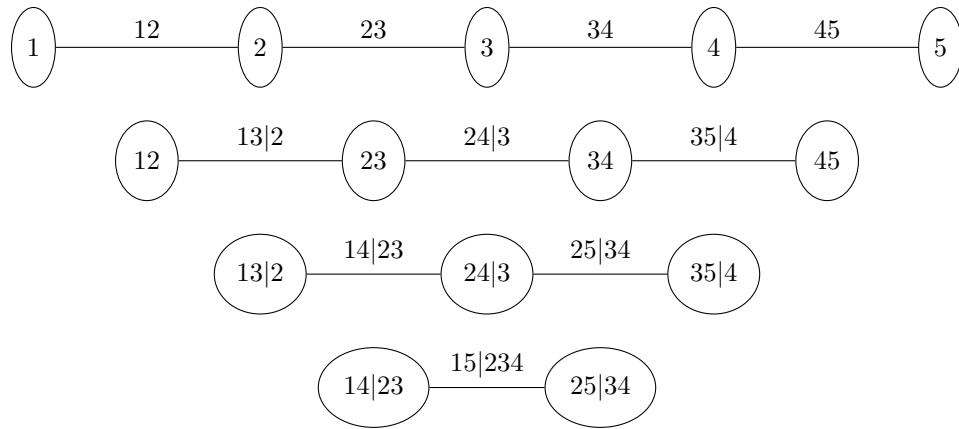
Figure 5.1: A 5-dimensional D-vine

The key feature of the D-vine is that the nodes of each tree only connect to adjacent nodes. The edges of the entire vine indicate the bivariate building blocks that contribute to the pair copula constructions. For data of T years, the joint distribution of $\mathbf{Y} = (Y_1, \ldots, Y_T)$ can be expressed based on D-vine as

$$f_{\mathbf{Y}}(y_1, \ldots, y_T) = \prod_{t=1}^{T} f_{Y_t}(y_t) \prod_{t=2}^{T} \prod_{s=1}^{t-1} \tilde{f}_{s,t|(s+1):(t-1)}(y_s, y_t | y_{s+1}, \ldots, y_{t-1}), \qquad (5.1)$$

where

$$\tilde{f}_{s,t|(s+1):(t-1)}(z_s, z_t | z_{s+1}, \ldots, z_{t-1}) = \frac{f_{s,t|(s+1):(t-1)}(z_s, z_t | z_{s+1}, \ldots, z_{t-1})}{f_{s|(s+1):(t-1)}(z_s | z_{s+1}, \ldots, z_{t-1}) f_{t|(s+1):(t-1)}(z_t | z_{s+1}, \ldots, z_{t-1})}$$

is ratio of the bivariate distribution to the product of marginals given the conditioning set, and can be viewed as the (conditional) "dependence ratio" with a ratio of one indicating conditional independence. The generic conditional quantities in (5.1) for mixed outcomes were derived as follows:

$$
\begin{aligned}
f_{Z|\boldsymbol{V}}(z|\boldsymbol{v}) &= f_{Z|V_h,\boldsymbol{V}_{\backslash h}}(z|v_h,\boldsymbol{v}_{\backslash h}) \\
&= \begin{cases}
\dfrac{C_{Z,V_h;\boldsymbol{V}_{\backslash h}}\left(F_{Z|\boldsymbol{V}_{\backslash h}}(0|\boldsymbol{v}_{\backslash h}),F_{V_h|\boldsymbol{V}_{\backslash h}}(0|\boldsymbol{v}_{\backslash h})\right)}{F_{V_h|\boldsymbol{V}_{\backslash h}}(0|\boldsymbol{v}_{\backslash h})} & z=0,v_h=0 \\[3mm]
\dfrac{f_{Z|\boldsymbol{V}_{\backslash h}}(z|\boldsymbol{v}_{\backslash h})c_{1,Z,V_h;\boldsymbol{V}_{\backslash h}}\left(F_{Z|\boldsymbol{V}_{\backslash h}}(z|\boldsymbol{v}_{\backslash h}),F_{V_h|\boldsymbol{V}_{\backslash h}}(0|\boldsymbol{v}_{\backslash h})\right)}{F_{V_h|\boldsymbol{V}_{\backslash h}}(0|\boldsymbol{v}_{\backslash h})} & z>0,v_h=0 \\[3mm]
c_{2,Z,V_h;\boldsymbol{V}_{\backslash h}}\left(F_{Z|\boldsymbol{V}_{\backslash h}}(0|\boldsymbol{v}_{\backslash h}),F_{V_h|\boldsymbol{V}_{\backslash h}}(v_h|\boldsymbol{v}_{\backslash h})\right) & z=0,v_h>0 \\[3mm]
f_{Z|\boldsymbol{V}_{\backslash h}}(z|\boldsymbol{v}_{\backslash h})c_{Z,V_h;\boldsymbol{V}_{\backslash h}}\left(F_{Z|\boldsymbol{V}_{\backslash h}}(z|\boldsymbol{v}_{\backslash h}),F_{V_h|\boldsymbol{V}_{\backslash h}}(v_h|\boldsymbol{v}_{\backslash h})\right) & z>0,v_h>0
\end{cases}
\end{aligned}
\tag{5.2}
$$

and

$$
\begin{aligned}
F_{Z|\boldsymbol{V}}(z|\boldsymbol{v}) &= F_{Z|V_h,\boldsymbol{V}_{\backslash h}}(z|v_h,\boldsymbol{v}_{\backslash h}) \\
&= \begin{cases}
\dfrac{C_{Z,V_h;\boldsymbol{V}_{\backslash h}}\left(F_{Z|\boldsymbol{V}_{\backslash h}}(0|\boldsymbol{v}_{\backslash h}),F_{V_h|\boldsymbol{V}_{\backslash h}}(0|\boldsymbol{v}_{\backslash h})\right)}{F_{V_h|\boldsymbol{V}_{\backslash h}}(0|\boldsymbol{v}_{\backslash h})} & z=0,v_h=0 \\[3mm]
\dfrac{C_{Z,V_h;\boldsymbol{V}_{\backslash h}}\left(F_{Z|\boldsymbol{V}_{\backslash h}}(z|\boldsymbol{v}_{\backslash h}),F_{V_h|\boldsymbol{V}_{\backslash h}}(0|\boldsymbol{v}_{\backslash h})\right)}{F_{V_h|\boldsymbol{V}_{\backslash h}}(0|\boldsymbol{v}_{\backslash h})} & z>0,v_h=0 \\[3mm]
c_{2,Z,V_h;\boldsymbol{V}_{\backslash h}}\left(F_{Z|\boldsymbol{V}_{\backslash h}}(0|\boldsymbol{v}_{\backslash h}),F_{V_h|\boldsymbol{V}_{\backslash h}}(v_h|\boldsymbol{v}_{\backslash h})\right) & z=0,v_h>0 \\[3mm]
c_{2,Z,V_h;\boldsymbol{V}_{\backslash h}}\left(F_{Z|\boldsymbol{V}_{\backslash h}}(z|\boldsymbol{v}_{\backslash h}),F_{V_h|\boldsymbol{V}_{\backslash h}}(v_h|\boldsymbol{v}_{\backslash h})\right) & z>0,v_h>0
\end{cases} \\
&= \begin{cases}
\dfrac{C_{Z,V_h;\boldsymbol{V}_{\backslash h}}\left(F_{Z|\boldsymbol{V}_{\backslash h}}(z|\boldsymbol{v}_{\backslash h}),F_{V_h|\boldsymbol{V}_{\backslash h}}(0|\boldsymbol{v}_{\backslash h})\right)}{F_{V_h|\boldsymbol{V}_{\backslash h}}(0|\boldsymbol{v}_{\backslash h})} & v_h=0 \\[3mm]
c_{2,Z,V_h;\boldsymbol{V}_{\backslash h}}\left(F_{Z|\boldsymbol{V}_{\backslash h}}(z|\boldsymbol{v}_{\backslash h}),F_{V_h|\boldsymbol{V}_{\backslash h}}(v_h|\boldsymbol{v}_{\backslash h})\right) & v_h>0,
\end{cases}
\end{aligned}
\tag{5.3}
$$

where $C_{Z,V_h;\boldsymbol{V}_{\backslash h}}(u_1,u_2)$ and $c_{Z,V_h;\boldsymbol{V}_{\backslash h}}(u_1,u_2)$ are the bivariate copula and density function associated with conditional distributions $F_{Z|\boldsymbol{V}_{\backslash h}}$ and $F_{V_h|\boldsymbol{V}_{\backslash h}}$, respectively. And $c_{k,Z,V_h;\boldsymbol{V}_{\backslash h}}(u_1,u_2) = \partial C_{Z,V_h;\boldsymbol{V}_{\backslash h}}(u_1,u_2)/\partial u_k$, for $k=1,2$.

Due to the parametric nature of the proposed model, we employ likelihood-based methods for estimation. We explored a sequential method that estimates and selects the bivariate copulas on a tree-by-tree basis. We start with the first tree, estimating the parameters and selecting the appropriate copulas from a given set of candidates. Fixing the parameters in the first tree, we then estimate the dependence parameters in the second tree for the candidate copulas and select the optimal. We continue estimating parameters and selecting copulas for the next tree of a higher order while holding the parameters fixed in all previous trees. We use a heuristic procedure based on a commonly used model selection method AIC to select the copula. The sequential approach reduces the number of models to compare extensively and thus helps to fast select an appropriate model for applied studies. The benefit could be substantial in the case of big data or high dimensional dependence.

The mixed D-vine provides a nature structure to derive the predictive distribution, not just a point prediction, of future claim cost. For i-th policyholder, denoting $\boldsymbol{Y}_i = (Y_{i1}, \cdots, Y_{iT})'$, the conditional distribution of $Y_{i,T+1}$ given $\boldsymbol{Y}_i$ is shown as:

$$f_{Y_{i,T+1}|\boldsymbol{Y}_i} = f_{i,T+1}(y) \prod_{t=2}^{T} \tilde{f}_{i,t,T+1|(t+1):T}(y_{it}, y|y_{i,t+1}, \cdots, y_{i,T}).$$

The derivation of the predictive distribution relies on the conditional independence assumption between Y_{i1} and $Y_{i,T+1}$ given $Y_{i2}, \cdots, Y_{i,T}$.

The proposed model captures the insurance data structure flexibly and enjoys the benefit of straightforward computation of predictive distribution. The data analysis emphasized the benefits of the mixed vine approach in both fitting the observations in the training sample and predicting the observations in the hold-out sample.

We are now working on another project which aims at experience rating using multivariate longitudinal claim data. For example, to predict future claim of auto insurance, we not only use claim history of auto insurance, but also incorporate the history information of homeowner insurance into our model.

5.2 Data Summary

In current project, we focus on joint modeling of building and contents (BC) and motor vehicle (MV) insurance of the LGPIF. Table 5.1 shows the total number of policies for each coverage type in the dataset for years $2006 - 2010$. Jointly, there are 2170 policies with both coverages.

Table 5.1: Empirical number of observations

	Total	0	1	2	3	4	5	>5
BC	5660	3,976	997	333	136	76	31	111
MV	2175	1,511	314	116	53	36	21	124
Joint	2170							

Potential rating variables, covariates, are displayed in Table 5.2. Here coverage and deductible are continuous covariates which is essential for copula identification.

Table 5.2: Description and summary statistics of covariates

Variable	Description	Mean
TypeCity	=1 if entity type is city	0.140
TypeCounty	=1 if entity type is county	0.058
TypeSchool	=1 if entity type is school	0.282
TypeTown	=1 if entity type is town	0.173
TypeVillage	=1 if entity type is village	0.237
TypeMisc	=1 if entity type is other	0.110
NoClaimCreditBC	=1 if no building and content claims	
	in prior year given BC coverage	0.328
NoClaimCreditMV	=1 if no motor vehicle claims	
	in prior year given MV coverage	0.054
lnCoverageBC	Coverage of BC line in logarithmic	2.119
	millions of dollars given BC coverage	(2.000)
lnCoverageMV	Coverage of MV line in logarithmic	-0.798
	millions of dollars given MV coverage	(1.626)
lnDeductBC	BC deductible level in logarithmic	7.155
	millions of dollars given BC coverage	(1.174)

Preliminary dependence measures for discrete claim frequencies can be obtained using simple correlation statistics such as polychoric correlations, which are the correlations of the latent normal variables, and Kendall's τ. Table 5.3 shows the correlations among the frequencies of the two coverages. Note that these dependence measures in Table 5.3 are

calculated before controlling for the effects of explanatory variables. The results in Table 5.3 should be taken with caution due to the following reasons. First, the polychoric correlation is negative, different from Kendall's τ and Spearman's ρ. The polychoric correlation is calculated based on likelihood. In our data, we have some large observations, and those observations are removed for numerical stability due to their almost 0 likelihood. Thus, the estimator is biased. Those large observations are essential for tail dependence, which is key to insurers' solvency. Second, as discussed in Chapter 1, the definitions of Kendall's τ and Spearman's ρ do not take the probability of ties into account. As a result, the correlation quantities are problematic for discrete outcomes. Third, the large values of the dependencies may be due to correlations in the covariates. We will further quantify the correlations using likelihood-based estimation after controlling the effects of covariates in Section 5.4.

Table 5.3: Correlation among frequencies of claims

Polychoric correlation	Kendall's τ	Spearman's ρ
-0.226	0.361	0.402

5.3 Marginal Models

We first analyze marginal models. From Table 5.1, it can be seen that the BC line contains a large number of zeros and a significant amount of ones. This motivates the usage of zero-one-inflated Poisson models described in Section 5.1.1. The distribution function can be expressed as

$$F_j(k|X_j, \beta_j) = \begin{cases} \pi_{j0} + (1 - \pi_{j0} - \pi_{j1})\exp(-\lambda_j) & k = 0, \\ \pi_{j0} + \pi_{j1} + (1 - \pi_{j0} - \pi_{j1})\sum_{i=0}^{k} \lambda_j^i \exp(-\lambda_j)\frac{1}{i!} & k > 0. \end{cases}$$

We fit both margins with four models: Poisson, negative binomial (NB), zero-inflated Poisson (zeroinflPoisson), and zero-one-inflated Poisson (zeroonePoisson) models. Note here for each margin, the univariate analysis is performed on the subset of the observations for

which the corresponding coverage amounts shown in Table 5.1 are positive. Thus, the sample size is 5,660 for BC and 2175 for MV.

To choose marginal models, as in Frees et al. (2016), the empirical and observed counts are compared. Table 5.4 shows the expected count for each frequency value under different models and the empirical values from the data for BC line. A Poisson distribution underestimates the zero proportions while zero-inflated and negative binomial models underestimate the proportion of 1 s. The zero-one inflated models do provide the best fits for simultaneously estimating the probability of zero and one. Chi-square goodness of fit statistics can be used to compare different models. Table 5.5 shows the result. It is calculated depending on Table 5.4. The zero-one-inflated Poisson model outperforms the other methods.

Table 5.4: Comparison between empirical values and expected values for BC line

	empical	zeroinflPoisson	zeroonePoisson	Poisson	NB
0	$3,976$	$4,038.125$	$3,975.388$	$3,709.985$	$4,075.368$
1	997	754.384	$1,024.226$	$1,012.267$	809.077
2	333	355.925	276.086	417.334	313.359
3	136	187.897	146.964	202.288	155.741
4	76	106.780	82.053	106.874	88.866
5	31	63.841	48.427	60.160	55.484
6	19	39.850	30.213	36.540	36.919
7	19	26.082	19.851	24.261	25.765
8	16	18.025	13.670	17.440	18.663
9	5	13.165	9.808	13.222	13.932
10	7	10.087	7.269	10.305	10.664
11	2	8.007	5.505	8.124	8.336
12	4	6.505	4.219	6.427	6.636
13	5	5.357	3.248	5.086	5.367
14	5	4.441	2.502	4.024	4.401
15	2	3.690	1.925	3.182	3.653
16	4	3.062	1.479	2.519	3.066
17	3	2.530	1.134	1.999	2.598
18	1	2.077	0.867	1.597	2.221
≥ 19	19	10.168	5.167	16.366	19.876
0 proportion	0.702	0.713	0.702	0.655	0.720
1 proportion	0.176	0.133	0.181	0.179	0.143

Similarly, Table 5.6 shows the expected count for each frequency value under different models and the empirical values from the data of MV line. Table 5.7 shows the goodness-of-fit tests result. It was calculated using the results in Table 5.6. The negative binomial

Table 5.5: Goodness-of-fit statistics for BC line

zeroinflPoisson	zeroonePoisson	Poisson	NB
154.573	77.056	105.201	88.086

model is selected, based on the test results.

Table 5.6: Comparison between empirical values and expected values for MV line

	empical	zeroinflPoisson	zeroonePoisson	Poisson	NB
0	$1,511$	$1,500.242$	$1,501.967$	$1,428.612$	$1,505.872$
1	314	274.685	304.915	341.613	313.934
2	116	116.722	99.257	118.070	114.897
3	53	64.507	59.515	64.434	60.342
4	36	45.843	41.992	47.616	38.704
5	21	37.612	33.736	39.559	27.622
6	22	32.141	28.799	33.402	20.898
7	11	27.075	24.741	27.547	16.352
8	13	21.977	20.771	21.873	13.060
9	15	17.052	16.817	16.649	10.573
10	9	12.621	13.066	12.143	8.639
11	12	8.917	9.728	8.499	7.108
12	9	6.025	6.947	5.721	5.879
13	7	3.901	4.764	3.712	4.884
14	6	2.426	3.143	2.325	4.071
15	5	1.451	1.998	1.407	3.405
16	1	0.837	1.225	0.824	2.854
17	3	0.465	0.726	0.467	2.399
18	3	0.250	0.416	0.256	2.020
≥ 19	8	0.252	0.479	0.271	10.489
0 proportion	0.695	0.690	0.691	0.657	0.693
1 proportion	0.144	0.126	0.140	0.157	0.144

Table 5.7: Goodness-of-fit statistics for MV line

zeroinflPoisson	zeroonePoisson	Poisson	NB
336.013	173.789	324.082	16.400

The coefficients for the selected models, zero-one-inflated Poisson for BC line and negative binomial for MV line, are in Table 5.8. We address that it is the benefit of employing copula regression models that the marginal models can be freely specified.

Table 5.8: Marginal coefficients

	Variable Name	BC (0-1 inflated Poisson)		MV (Negative Binomial)	
		Coef.	s.e.	Coef.	s.e.
Count	(Intercept)	-1.540	0.125	-0.929	0.109
	lnCoverage	0.751	0.023	0.708	0.036
	lnDeduct	-0.020	0.017		
	NoClaimCredit	-0.395	0.131	-0.370	0.146
	TypeCity	-0.143	0.079	0.231	0.149
	TypeCounty	-0.250	0.087	1.518	0.132
	TypeMisc	-0.195	0.179	-0.352	0.301
	TypeSchool	-1.157	0.085	0.651	0.131
	TypeTown	0.186	0.175	-1.085	0.244
	size			1.427	
Zero	(Intercept)	-4.755	0.448		
	lnCoverage	-0.580	0.078		
	lnDeduct	0.879	0.062		
	NoClaimCredit	0.536	0.280		
One	(Intercept)	-5.533	0.639		
	lnCoverage	-0.047	0.094		
	lnDeduct	0.577	0.084		
	NoClaimCredit	0.300	0.353		

5.4 Copula Identification

Given well-fitting marginal models, now we are in a position to conduct the dependence analysis. We focus on the 2170 policies with both BC and MV coverages. The nonparametric estimator is fit with bandwidth selected by the process explored in Section 4.4. The fitted nonparametric copulas are displayed in Figure 5.2 as the solid curves.

To address the practical issue of parametric copula selection, we compare the nonparametric estimator with different commonly used parametric copulas fit through MLE. Table 5.9 includes the parameters of different copulas. When the parameters are transformed to Kendall's τ, it is not surprising that the dependence is weaker than the raw dependence from Table 5.3 that was computed before introducing covariates. Figure 5.2 shows the graphical comparisons between different parametric copulas with the nonparametric estimator. As in Section 4.3, it is difficult to distinguish among different copulas when the dependence is weak. From Figure 5.2, we are only able to conclude that the Clayton copula does not fit well.

We further summarize the discrepancies numerically using the distance defined in (4.2)
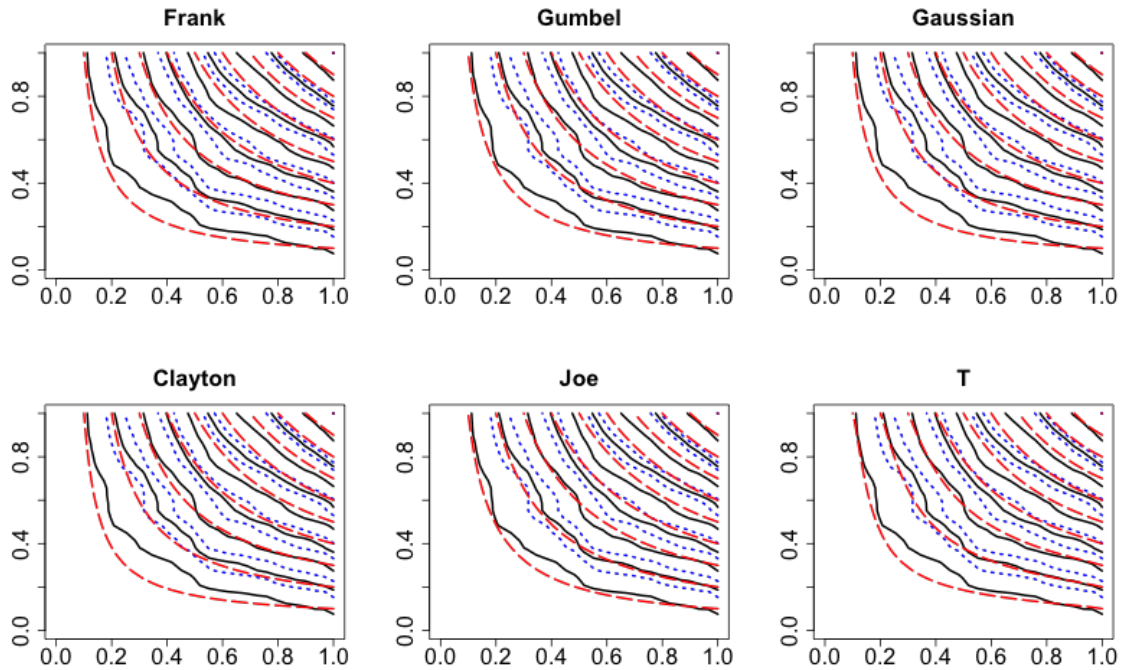
Figure 5.2: Plot of $\hat{C}(s, t; \hat{\beta})$ (solid) and confidence intervals (dotted) compared with parametric copulas contours (dashed)

Table 5.9: Parameters from different parametric copulas

	Estimate	s.e.	Kendall's τ
Gumbel	1.040	0.015	0.038
Joe	1.042	0.018	0.024
T(df=4)	0.072	0.038	0.046
Frank	0.718	0.221	0.079
Gaussian	0.125	0.033	0.079
Clayton	0.223	0.075	0.100

in Table 5.10. The Frank, Gaussian, and Clayton copulas can be excluded due to their large discrepancies. The performance of the t, Gumbel, and Joe copulas seem similar, which suggests that there is upper tail dependence in this dataset. To take the uncertainty into account, we do bootstrap with the number of replication as 500 to obtain the standard errors of the distances. Since the standard errors are comparable, given the smallest mean distance in Table 5.10, the Joe copula seems to best describe the dependence.

Table 5.10: Distances $d(\hat{C}(\cdot; \hat{\beta}), \tilde{C}_{\hat{\theta}})$ of different parametric copulas with the nonparametric estimator(multiplied by 1000)

	Gumbel	Joe	t	Frank	Gaussian	Clayton
Estimate	0.633	0.635	0.646	0.711	0.701	0.885
s.e.	0.240	0.251	0.240	0.239	0.237	0.265

Since the distances of parametric copulas with our nonparametric estimator may not be normally distributed, standard errors may not be informative enough to quantify the uncertainty. Figure 5.3 displays the distribution of the distances of different copula families constructed from bootstrap samples. The Joe, Gumbel, and t copulas appear better than the rest in the sense that their distances are mostly distributed around small values.
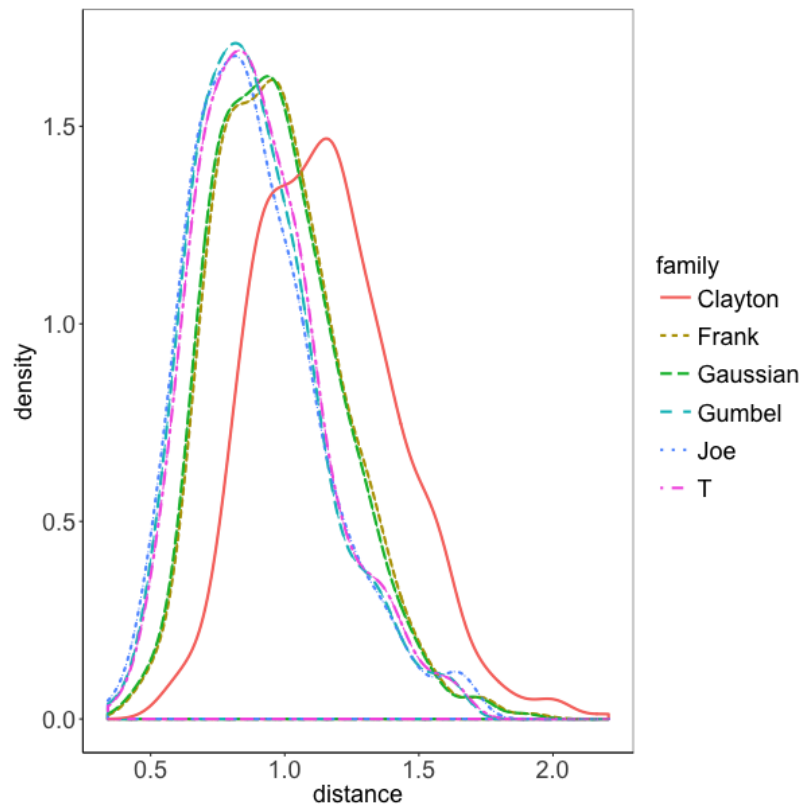
Figure 5.3: Density plot of distances (multiplied by 1000) from different parametric copulas

Chapter 6

Copula Identification for Mixed (Tweedie) Outcomes

In this chapter, we extend the framework of copula identification to mixed type of variables, i.e., random variables with both discrete and continuous components. Due to the discrete component, copula identification has been a problem.

Following the flow of the main body of the dissertation, we organize this chapter as following. Section 6.1 lays out the problem. Section 6.2 summarizes the commonly used marginal models and proposes the perturbed empirical residual distribution function for mixed data as foundation for copula identification. In Section 6.3, we present the nonparametric estimator of copulas. Section 6.4 includes a simulation study to explore the behavior of the estimator under different scenarios. This project is ongoing; future work includes more thorough simulations, data analysis, and theoretical proofs.

6.1 Introduction

In insurance applications, many claim data come in a multivariate fashion. For example, commercial insurance companies might provide their policyholders both motor vehicle coverage and building and contents coverage, and it is natural for insurers to track claims by

coverage types. When an insurer has a collection of multivariate risks, understanding the dependencies is essential for estimating the distribution of the portfolio, which is important to firm solvency and profitability.

It is not an easy task to model multivariate outcomes when their marginal distributions are complicated. It is common that insurance claim outcomes follow a mixed distribution of a point mass at zero with a positive distribution, for which zero corresponds to no claims and the positive distribution describes the amount of claims given occurrence. Hence, the multivariate normal distribution can not capture the mixed feature of claim data. Meanwhile, various characteristics of the policyholders are typically supplied to the insurer. As an example, in motor vehicle insurance, rating variables include the age and sex of the driver, type and age of the vehicle, and so forth. These variables are important for ratemaking.

Marginally, to accommodate the special feature of claim data and make use of covariates meanwhile, two types of models are prominently applied. The first method is Tweedie GLMs (Ohlsson and Johansson 2006). The latter is the frequency-severity approach (Frees et al. 2013), for which the frequency of claim and the severity given occurrence are modeled separately.

To capture the dependence structure, one natural way is compound Poisson linear mixed models, for which outcomes are correlated through random effects (McCulloch and Neuhaus 2001). Though, the generalized linear mixed model is computationally demanding (Zhang 2013). In addition, the model does not maintain the intuitive Tweedie marginal distributions after integrating the random effects.

In recent years, copulas have been widely used to study dependencies in many areas including, but not limited to, insurance (Frees and Valdez (1998)), finance (Li 1999) and survival analysis (Shih and Louis 1995); see Nelsen (2007) for an introduction. As a useful tool to separate the investigation of margins and dependence structure, a copula model preserves the body of work established for marginal models. There are few applications of copulas on mixed type of data in the literature. Frees et al. (2016) studied dependence of frequency and severity parts separately, in which each part is fit with a separate copula.

Thus, methodologies of copulas on continuous and discrete outcomes in the literature can be readily applied; see a thorough summarization in Joe (2014).

We focus on the parsimonious models for which a single copula is built on the aggregate loss from different types of insurance coverages for its easy interpretation. Shi (2016) modeled claims from different types of coverage in automobile insurance using copula-based multivariate Tweedie models. Shi and Yang (2017) modeled temporal dependence of claims by building vine copulas (Aas et al. 2009; Panagiotelis et al. 2012) while fit the margins with two-part models.

Yet, there are issues when applying copulas on mixed type data. It has been addressed in the literature that copula identifiability is an issue for the discrete outcomes (Genest and Nešlehová 2007). Sklar showed the uniqueness of copulas is only guaranteed at the range of marginal distribution functions. In the i.i.d. case, denote the probability of zero as p_0, then the range of distribution function for mixed type of data is $\{0\} \cup [p_0, 1]$. From Sklar's theorem, the copula is not unique in the range $(0, p_0)$. Hence the identifiability remains a problem for mixed type of data due to the discrete component at zero. When we do regression as under our sampling scheme, p_0 changes with the covariates. If we further assume the copula does not change with the covariates, the range for copula identifiability widens to $\{0\} \cup [p_b, 1]$, where p_b is the lower bound of the support of p_0. Hence, when p_0 distributes at wide range, identifiability is less of a concern.

In addition, how to correctly specify a copula model has remained a question for mixed type of data. Copula identification begins after marginal models have been fit. Under continuity, we use the "Cox-Snell" residuals (Cox and Snell 1968) from marginal models to check for association. We can create scatter plots of the residuals (pp and tail dependence plots) (Joe 2014) or through formal tests (Li and Genton 2013) to look for dependence structures and identify a parametric copula.

However, the Cox-Snell residuals are not uniformly distributed for mixed type data. Hence the joint distribution of multivariate Cox-Snell residuals is not as informative. In practice, overall goodness-of-fit statistics, such as AIC, BIC, and likelihood, are used to

choose the best model among candidates (Shi and Yang 2017). Vuong's test (Vuong 1989) can be applied to further compare if the models are statistically significantly different. However, these methods are not diagnostic for adequacy of fit and do not suggest improvements.

To identify the dependence structure for mixed type of data, in this dissertation, we develop a nonparametric copula estimator. Instead of building the estimator on uniformly distributed residuals, we construct the estimator based on a subset of the observations. For practitioners who prefer to use parametric copulas, the proposed nonparametric estimator can also serve as a diagnostic and specification tool for selecting a parametric form of copulas. Adequacy of fit can be checked by comparing the fitted parametric copula with the nonparametric estimator.

6.2   Univariate

6.2.1   Marginal Models

The Tweedie distribution is defined as a Poisson sum of gamma random variables. Specifically, suppose that N has a Poisson distribution with mean λ, representing the frequency of claims. Let Z_j be an i.i.d. sequence, independent of N, with each Z_j having a gamma distribution with shape parameter α and rate parameter γ, representing the amount of a claim. Then, $Y = Z_1 + \ldots + Z_N$ is a Poisson sum of gammas.

To understand the mixture aspect of the Tweedie distribution, first note that it is straightforward to compute the probability of zero as $\Pr(Y = 0) = \Pr(N = 0) = e^{-\lambda}$. The distribution function can be computed using conditional expectations,

$$\Pr(Y \leq y) = e^{-\lambda} + \sum_{n=1}^{\infty} \Pr(N = n) \Pr(Y \leq y), \qquad y \geq 0.$$

Because the sum of i.i.d. gammas is a gamma, $Z_1 + \ldots + Z_n$ has a gamma distribution with

parameters $n\alpha$ and γ. For $y > 0$, the density of the Tweedie distribution is

$$f_Y(y) = \sum_{n=1}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \frac{\beta^{n\alpha}}{\Gamma(n\alpha)} y^{n\alpha-1} e^{-y\beta}.$$

We address that the density function is not analytically tractable, hence numerical procedures are required for fitting Tweedie distribution.

From this, straightforward calculations show that the Tweedie distribution is a member of the linear exponential family. Now, define a new set of parameters μ, ϕ, P through the relations

$$\lambda = \frac{\mu^{2-P}}{\phi(2-P)}, \qquad \alpha = \frac{2-P}{P-1} \qquad \text{and} \qquad \frac{1}{\beta} = \phi(P-1)\mu^{P-1}.$$

Easy calculations show that

$$\text{E } Y = \mu \qquad \text{and} \qquad \text{Var } Y = \phi\mu^P,$$

where $1 < P < 2$. The Tweedie distribution can also be viewed as a choice that is intermediate between the Poisson and the gamma distributions.

To perform regression, one can relate the mean with covariates through a link function,

$$\eta_\mu(\mu_i) = \mathbf{x}_i'\beta.$$

In the basic form of the Tweedie regression model, the scale (or dispersion) parameter ϕ is constant. However, if one begins with the frequency-severity structure, calculations show that ϕ depends on the risk characteristics (*cf.*, Frees et al. 2014). Because of this and the varying dispersion (heteroscedasticity) displayed by many datasets, researchers have devised ways of accommodating and/or estimating this structure. The most common way is the so-called "double GLM" procedure proposed in Smyth (1989) that models the dispersion as a known function of a linear combination of covariates (as for the mean, hence the name "double GLM"), i.e.,

$$g_\phi(\phi_i) = \mathbf{z}_i'\eta.$$

Though, the inherent tail property of Tweedie GLMs may not be enough to accommodate the long-tail and right skewed feature in some applications, especially in commercial insurance. For example, Frees et al. (2016) demonstrated the Tweedie GLM provides poor fitting for claim of building and contents insurance of the LGPIF, from both in-sample and out-of-sample validation.

Alternatively, frequency-severity models, or two-part models provide a flexible way to fit claim data. Specifically, the frequency part, indicating whether or not a claim has occurred or the number of claims, and severity part, indicating the amount of a claim given occurrence, are modeled separately. In Frees et al. (2016), the frequency of building and contents claims is modeled by a zero-one-inflated Poisson model while the severity part follows a *GB2* distribution. The zero-one-inflated Poisson model captures the feature that not only a large point mass zero, but also significant proportion of ones are observed for the frequency of building and contents claims. The *GB2* distribution can fit heavy tail distributions well. In addition, instead of assuming independence between frequency and severity, the dependence between frequency and severity can be further modeled (Czado et al. 2012; Krämer et al. 2013). Though one needs to do variable selection for frequency and severity part separately and more parameters need to be fit for the two-part model. In contrast, Tweedie GLMs are more parsimonious.

6.2.2 Perturbed Probability Integral Transformation

Given a marginal model, as mentioned in Chapter 2, the following step in copula regression is to transform the data into uniform pseudo observations. For a variable Y, if Y is continuous, then $F(Y)$ is known as the probability integral transform, and is uniformly distributed as in (2.1). Note here F might depend on covariates X. Let $(X_i, Y_i), i = 1, \ldots, n$ be an i.i.d sample of (X, Y). With fitted marginal model $\hat{F}$, by plugging the data into the formula of the probability integral transform, we get a sequence of Cox-Snell residuals (Cox and Snell 1968) $\hat{F}_i(Y_i), i = 1, \ldots, n$, which should present uniform trend under the true model.

However, the Cox-Snell residuals are not uniformly distributed for mixed type data when

there are many zeros. Figure 6.1 displays the Cox-Snell residuals of a simulated Tweedie example, which do not present uniform trend under the true model due to the zero point mass. Therefore, we build an alternative of the Cox-Snell residuals for mixed data.
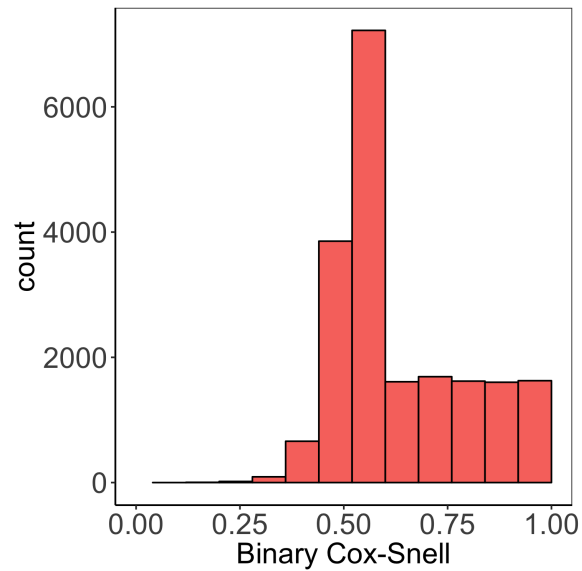
Figure 6.1: Histogram of Cox-Snell residuals of mixed data

To analyze the behavior of probability integral transforms for mixed outcomes, we use a generalized notation for the distribution function, which applies to both the Tweedie GLMs and the two-part models. Denote p_0 as zero probability and g as density for positive part. Hence the density of Y is

$$
f(y) = \begin{cases} p_0 & y = 0 \\ g(y) & y > 0. \end{cases}
$$

The distribution function is

$$
F(y) = \begin{cases} p_0 & y = 0 \\ p_0 + (1 - p_0)G(y) & y > 0. \end{cases}
$$

For a fixed point $s \in (0, 1)$, the distribution function of the variable $F(Y)$ is that

$$
P(F(Y) \leq s) = \begin{cases} 0 & p_0 > s \\ P(Y = 0) + P\left(0 < Y \leq G^{-1}\left(\frac{s - p_0}{1 - p_0}\right)\right) & p_0 \leq s \end{cases}
$$
$$
= \begin{cases} 0 & p_0 > s \\ = p_0 + (1 - p_0)G\left(G^{-1}\left(\frac{s - p_0}{1 - p_0}\right)\right) = s & p_0 \leq s \end{cases} \tag{6.1}
$$

That is, the probability integral transforms are not uniform when $p_0 > s$.

Now we construct an alternative to the empirical distribution of probability integral transformations $\hat{U}(\cdot)$ as in (2.2). We first consider this problem treating model F as known and will replace it by its estimator later. Suppose we have the data $(X_i, Y_i), i = 1, \ldots, n$ independently identically distributed as (X, Y) and a fixed point s. From (6.1), instead of plugging all the observations into (2.2), we only focus on a subset of the observations for which $p_{0i} \leq s$. Hence, the perturbed empirical residual distribution function for mixed data is

$$
\hat{U}(s) = \frac{\sum_{i=1}^{n} 1(F_i(Y_i) \leq s, p_{0i} \leq s)}{\sum_{i=1}^{n} 1(p_{0i} \leq s)},
$$

recall F_i is related to covariates.

We can further smooth the estimator. Instead of using indicator for $p_{0i} \leq s$, we can adapt the smoothed empirical distribution, i.e. $K((s - p_{0i})/\epsilon_n)$, where K is a distribution function of a kernel function such as Epanechnikov distribution

$$
K(u) = \begin{cases} 0 & u < -1 \\ \frac{1}{4}(2 + 3u - u^3) & |u| \leq 1 \\ 1 & u > 1. \end{cases}
$$

Note that when $s < p_{0i} \leq s + \epsilon_n$, $P(F_i(Y_i) \leq s) = 0$, though $P(Y_i = 0) = p_{0i} \approx s$. Hence

we adjust for these observations. Denote

$$\tilde{Y}_i = \begin{cases} 1(F_i(Y_i) \leq s) & s \geq p_{0i} \\ 1(Y_i = 0) & s < p_{0i}. \end{cases}$$

Thus, the smoothed perturbed empirical residual distribution function is

$$\hat{U}(s) = \sum_{i=1}^{n} W_{ni}(s)\tilde{Y}_i, \tag{6.2}$$

where

$$W_{ni}(s) = \frac{K((s - p_{0i})/\epsilon_n)}{\sum_{i=1}^{n} K((s - p_{0i})/\epsilon_n)}.$$

The term $\hat{U}(s)$ should be close to s for $s \in (0,1)$ under true model.

6.3 Copula Estimation

Extend our notations to two dimensions, now we have bivariate outcomes (Y_1, Y_2) with corresponding covariates (X_1, X_2), marginal distribution functions F_1 and F_2, and marginal zero probabilities p_{10}, p_{20}.

6.3.1 Parametric Copula Estimation

In practice, the copula can be estimated through MLE with selected copula parametric family. To illustrate the general principles, consider the bivariate case. Suppressing the i index and covariate notation, the joint distribution is

$$f(y_1, y_2) = \begin{cases} C(p_{10}, p_{20}) & y_1 = 0, y_2 = 0 \\ f(y_1)C_1(F_1(y_1), p_{20}) & y_1 > 0, y_2 = 0 \\ f(y_2)C_2(p_{10}, F_2(y_2)) & y_1 = 0, y_2 > 0 \\ f(y_1)f(y_2)c(F_1(y_1), F_2(y_2)) & y_1 > 0, y_2 > 0 \end{cases}$$

where C_j denotes the partial derivative of the copula with respect to jth component.

6.3.2 Copula Estimation

Under continuity, there is an unique copula related to (Y_1, Y_2), which is the joint distribution function of the marginal probability integral transforms, as in (3.2).

For mixed outcomes, as an extension to (6.1), we can see that only when $s \geq p_1, t \geq p_{i2}$,

$$P(F_1(Y_1) \leq s, F(Y_2) \leq t) = F(F_1^{-1}(s), F_2^{-2}(t)) = C(s, t).$$ Hence when we have a sample, to build copula estimator, we focus on the subset of the observations for which $s \geq p_{i10}, t \geq p_{i20}$, and the intuitive copula estimator is

$$\hat{C}(s, t) = \frac{\sum_{i=1}^n 1(F_1(Y_{i1}) \leq s, F_2(Y_{i2}) \leq t, p_{i10} \leq s, p_{i20} \leq t)}{\sum_{i=1}^n 1(p_{i10} \leq s, p_{i20} \leq t)}. \tag{6.3}$$

Similar to the univariate case, we can further smooth the estimator. Define

$$\tilde{Y}_i = \begin{cases} 1(F_1(Y_{i1}) \leq s, F_2(Y_{i2}) & p_{i10} \leq s, p_{i20} \leq t \\ 1(Y_{i1} = 0, F(Y_{i2}|x, \beta) \leq t) & p_{i10} > s, p_{i20} \leq t \\ 1(F(Y_{i1}|x, \beta) \leq s, Y_{i2} = 0) & p_{i10} \leq s, p_{i20} > t \\ 1(Y_{i1} = 0, Y_{i2} = 0) & p_{i10} > s, p_{i20} > t. \end{cases}$$

Then the smoothed copula estimator is

$$\hat{C}(s, t) = \sum_{i=1}^n W_{ni}(s; x_i, \beta) \tilde{Y}_i, \tag{6.4}$$

where

$$W_{ni}(s; x_i, \beta) = \frac{K((s - p_{i1})/\epsilon_n) K((t - p_{i2})/\epsilon_n)}{\sum_{i=1}^n K((s - p_{i1})/\epsilon_n) K((t - p_{i2})/\epsilon_n)},$$

and K is a distribution function on a bounded support.

6.4   Simulation

In this study, we consider 5000 policyholders, each has insurance coverage in two lines of business, whose expected claims are are based on the function

$$E(Y_j|X_j) = \exp(\beta_{j0} + X_j\beta_{j1}),$$

where each X_j is generated from a standard normal distribution independently. Since the distribution of marginal zero probabilities plays an essential role for copula identification, here we do three scenarios on the range of zero probabilities by controlling the value of β_{j0}. Throughout the scenarios, we assume the dispersion parameter $\Phi = 10$ and $P = 1.6$.

- Many zeros. $\beta_{10} = 0, \beta_{11} = 2$, $\beta_{20} = 0, \beta_{21} = 1.5$. Average $p_{10} = 0.754$, $p_{20} = 0.779$.

- Moderate zeros. $\beta_{10} = 2, \beta_{11} = 2$, $\beta_{20} = 2, \beta_{21} = 1.5$. Average $p_{10} = 0.546$, $p_{20} = 0.574$.

- Few zeros. $\beta_{10} = 4, \beta_{11} = 2$, $\beta_{20} = 4, \beta_{21} = 1.5$. Average $p_{10} = 0.281$, $p_{20} = 0.293$.

Figure 6.2 displays the histogram of zero probabilities and corresponding contour plots of the copula estimator. The top row corresponds to the many zeros scenarios. From the histogram, p_{10} and p_{20} are mostly greater than 0.7, as a result, we could identify the copula at the range $[0.7, 1] \times [0.7, 1]$ but not other area. Similarly, when most p_{10} and p_{20} are greater than 0.4, we are able to identify the copula at $[0.4, 1] \times [0.4, 1]$. In the case when the zero probabilities are small, the copula can be identified.
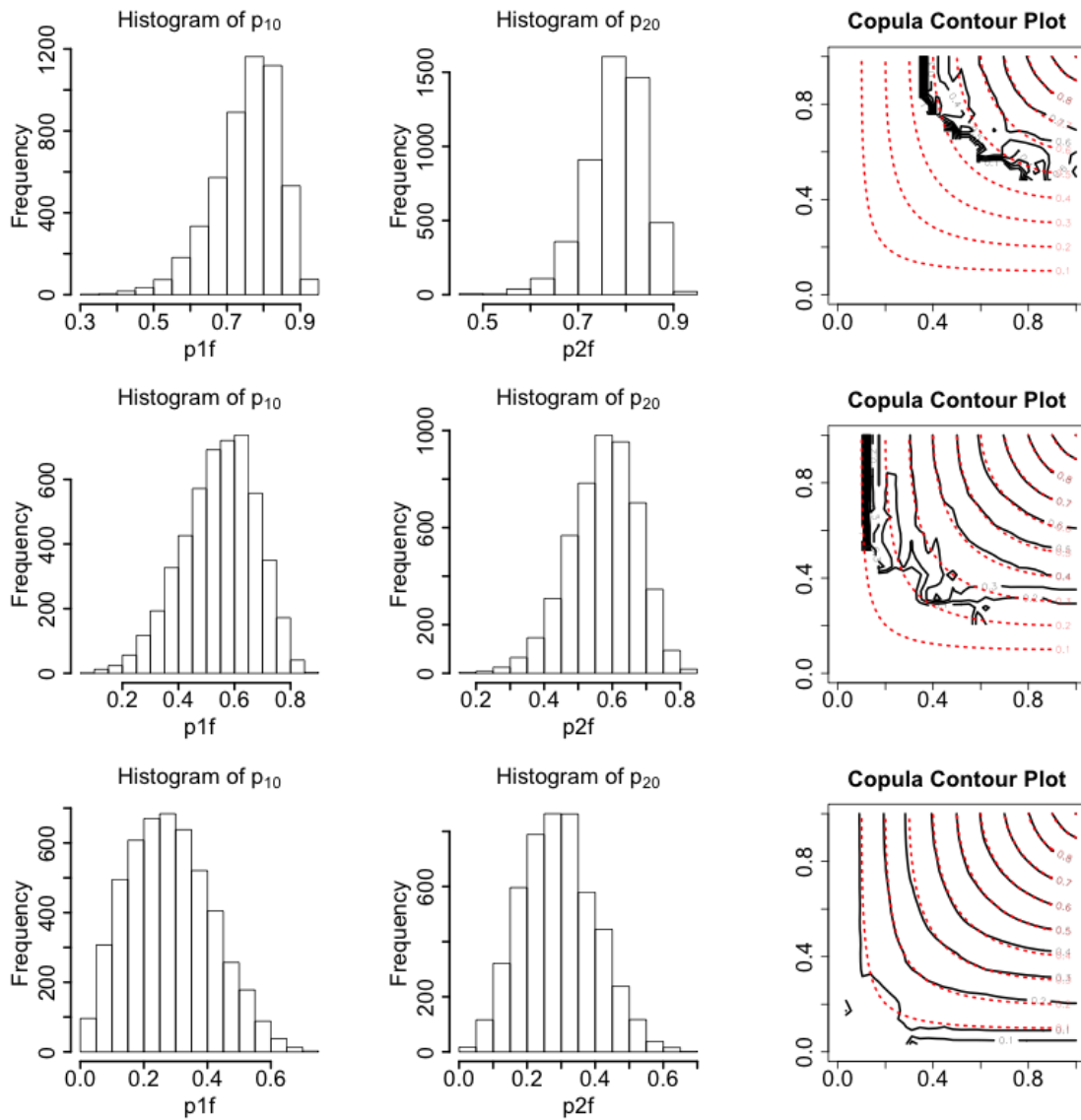
Figure 6.2: Histogram of p_0 and corresponding contour plots of the copula estimator. Top row: many zeros. Middle row: moderate zeros. Bottom row: few zeros.

Chapter 7

Concluding Remarks and Ongoing Work

7.1 Conclusions

In this dissertation, we considered modeling multivariate discrete outcomes with copulas. We explored the dependence modeling in the practical regression settings. Our main contribution is the proposal of a nonparametric copula estimator to specify the dependence structure under discreteness when the premises of methodologies under continuity are violated. We also showed its asymptotic properties. Using a simulation study, we concluded that first, the estimator behaves better with small bias and variance when the marginal means are large, which is consistent with the theoretical results. Second, when used as a diagnostic tool, the nonparametric estimator can exclude false models easily when the dependence is high and the discreteness level is low. The data analysis suggested in the LGPIF dataset, there is upper tail dependence between the frequencies from building and contents coverage and motor vehicle coverage.

We acknowledge that many potential improvements can be made for our study. First, the estimator can be further refined. In this dissertation, we applied the local average approach. Local polynomial estimators can be explored to reduce bias on boundary. In addition, the

bandwidth selector we proposed chooses a global bandwidth. Since for our estimator, we have more observations at the right upper corner than the lower left corner, the variable bandwidth in Fan and Gijbels (1995) might be applied. Second, recall that in Chapters 4 and 5, we used the proposed nonparametric copula estimator as the foundation for goodness-of-fit tests of parametric copulas. We computed the distance of different parametric copulas with our nonparametric estimator $d(\hat{C}(\cdot); \hat{\beta}, \tilde{C}_{\hat{\theta}})$ in (4.2) and chose the model with smallest distance. We will formalize its usage by deriving the asymptotic properties of the distance using techniques for empirical processes. These are areas for our future work.

Below, I describe related ongoing research projects. Firstly, as mentioned in Chapter 2, the univariate proposed perturbed empirical residual distribution function can be used as a tool for goodness-of-fit test for marginal models. We formalize the idea in Section 7.2. Finally, in Section 7.3, we propose an alternative way of doing diagnosis for copula models using a transformation proposed by Ruschendorf (Rüschendorf 2009).

7.2 Univariate Model Diagnosis

Due to the wide application of univariate regression models on discrete outcomes, model specification and diagnosis have been extensively studied. In practice, information criteria such as AIC and BIC can be applied to choose the best model among candidates. Vuong's tests described in Section 5.1.1 provides an alternative way of statistically differentiating two parametric models which could be non-nested. Though, information criteria and Vuong's tests do not provide diagnosis for adequacy of fitting. As a result, the preferred model can not be adopted with full confidence.

As discussed in Section 2.2, the perturbed empirical residual distribution function defined for discrete outcomes $\hat{U}(\cdot; \hat{\beta})$ as in (2.4), which is an approximation to the empirical distribution function of Cox-Snell residuals under continuity, should be close to the identity function under correct model specification. Thus, it can be used to check goodness-of-fit

for marginal models on discrete outcomes when there are continuous covariates.

Alternatively, there is substantial literature on appropriate residuals for checking goodness-of-fit (Cook and Weisberg 1982; McCullagh and Nelder 1989). The aim was to construct residuals that are approximately normally distributed under true models, hence the discrepancy suggests the inadequacy of fitting. The two most commonly used ones are Pearson and deviance residuals. Pearson residuals generalize the idea of residuals in linear models, and are of the form

$$r_P(y_i; \theta) = \{y_i - \mathrm{E}_\theta(y_i)\} / \mathrm{SD}_\theta(y_i),$$

where θ is the parameter, and $\mathrm{E}_\theta(y)$ and $\mathrm{SD}_\theta(y)$ are the expected value and variance when the model parameter is θ, respectively. Deviance residuals are based on the contribution of each point to the likelihood and are the signed difference between the likelihood of the fitted model and the saturated model for which the number of parameters is same as the number of observations, i.e.,

$$r_D(y_i; \theta) = \mathrm{sgn}(y_i - \mathrm{E}_\theta(y_i)) \left\{ 2 \left[l(y_i; \tilde{\theta}) - l(y_i; \theta) \right] \right\}^{1/2},$$

where $l(y_i; \theta)$ denotes the likelihood taking value at y_i when the parameter is set to be θ, and $\tilde{\theta}$ is the parameter of the saturated model.

Through a simulation study, we explore model diagnosis using our perturbed empirical residual distribution function as well as Pearson and deviance residuals. Two aspects will be considered: the proximity to normality (or identity for the proposed method) under true models and the discrepancy with null pattern under misspecified models.

7.2.1 Simulation Study

In this study, we consider sample size $n = 1000$. Here we study three commonly used GLMs: Poisson with log link, negative binomial with log link and logistic regression for binary outcomes. The location parameter $\mu = X'\beta = \beta_0 + X_1\beta_1$, where $X_1 \sim N(0,1)$.

As discussed by Pierce and Schafer (1986), many residuals for GLM require large expected value to achieve normal approximation, which was referred as m-asymptotics. This is in line with the fact that the level of discreteness plays an important role for copula identification as in Chapter 3. Therefore, here we do simulation on three levels of discreteness for Poisson and negative binomial models:

- Small mean: $\beta_0 = -2, \beta_1 = 2$

- Medium mean: $\beta_0 = 0, \beta_1 = 2$

- Large mean: $\beta_0 = 5, \beta_1 = 2$

The dispersion parameter for negative binomial is set to be 0.7.

For each of the experiment, we generate the data, fit the model, and calculate the residuals or perturbed residuals $\hat{U}(\cdot; \hat{\beta})$. Then we summarize the results graphically by providing the curve of $\hat{U}(\cdot; \hat{\beta})$ and the QQ-plots of the other two types of residuals. Recall that given correct model specification, the residuals are expected to be normally distributed, and $\hat{U}(\cdot; \hat{\beta})$ is expected to be the identity function. Thus, the null pattern is along the diagonal.

Closeness to Null Patterns under Correct Model Specifications. Figure 7.1 summarizes the result for Poisson outcomes graphically. Here we generate the data using Poisson GLM, fit the correct model, and then calculate the residuals. The upper row corresponds to the small mean scenario. As anticipated, deviance and Pearson residuals are far apart from normality due to small m, while the proposed perturbed empirical residual distribution function is close to the identity function. That is, our method provides more reliable conclusion for cases with high level of discreteness. When we move to the middle row corresponding to the medium mean level, our method keeps the pattern along the diagonal, and the deviance residuals are getting closer to be normally distributed, while the Pearson residuals still have large discrepancy with normality. As the mean increases to large case (lower row of Figure 7.1), all three methods appear close to the null pattern.

We also checked the results of the negative binomial regression model under different mean levels, which is not reported here. In general, the conclusion is consistent with Figure 7.1 that deviance and Pearson residuals are more vulnerable to high level of discreteness in the data.
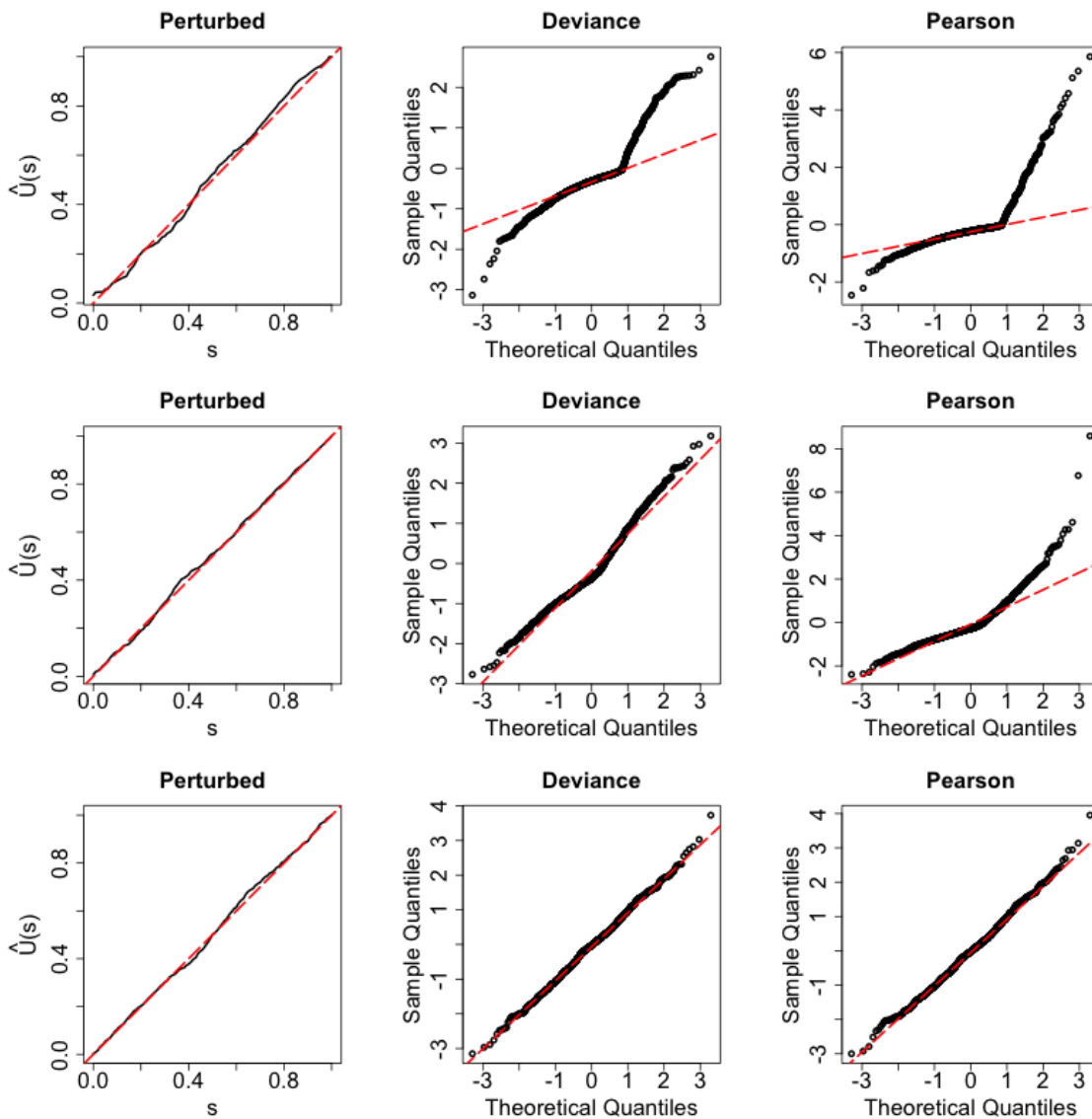

Figure 7.1: Diagnostic plots for Poisson outcomes. The three rows correspond to small, medium, and large mean levels.

Figure 7.2 shows the results for binary outcomes. Here we only present the result when $\beta_0 = 0$ since binary outcomes are highly discrete for all the settings. We can see that all

three methods have large discrepancy. As mentioned in McCullagh and Nelder (1989), the residual plots for binary outcomes are not as informative. Hence diagnosis for binary models is not our concern here.
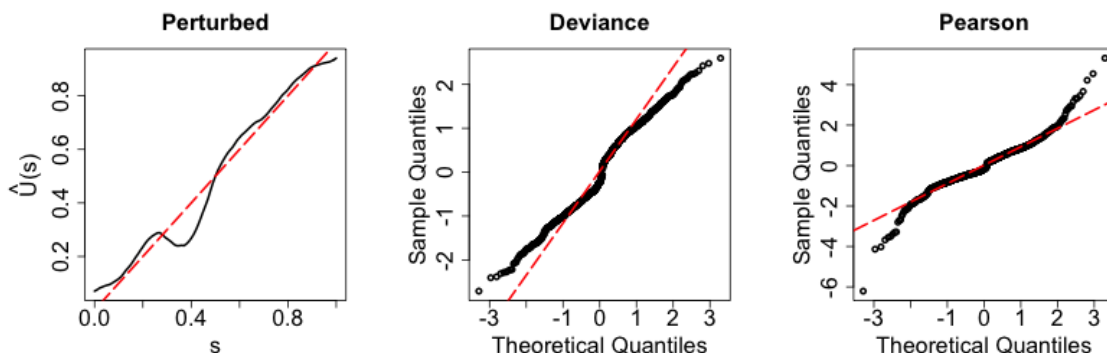

Figure 7.2: Diagnostic plots for binary outcomes.

Discrepancy under Misspecification. As concluded in the previous section, the proposed method is closer to null pattern under true model compared with the residual based methods. Now we explore the other side of the problem, the power of the methods for detecting misspecification.

For discrete outcomes, one of the most common problems is overdispersion. In the following example, we simulate the data with the negative binomial model and compare the residuals from the true model and a Poisson model, and thus inadequacy of fitting is expected from Poisson model.

Figure 7.3 summarizes the results from the proposed method and the residual-based methods under small mean scenario graphically. The first row shows the results of true model and the second row shows those when the overdispersion issue is not taken care of. We can see that the deviance and Pearson residuals are less informative, since they present comparably large discrepancy whether the model is correctly specified or not. In contrast, from the behavior of the perturbed empirical residual distribution function, we see significantly larger discrepancy when the models are misspecified. Thus, the better model can be picked up. It is noticeable that the curve of $\hat{U}(\cdot; \hat{\beta})$ seem unstable at lower

left corner, compared with Figure 7.1. This is due to small sample size at lower tail part, which was further explained in Section 3.3. When constructing formal goodness-of-fit tests, we can diminish the influence of this part by downweighting the corresponding area. As the mean increases to the medium and large levels, whose results are not reported here, deviance residuals become more informative, while Pearson residuals still show comparable discrepancies when the model is correctly or incorrectly specified. Though these simulations are not extensive enough. Simulations with different sample sizes and more replications will be our future work.
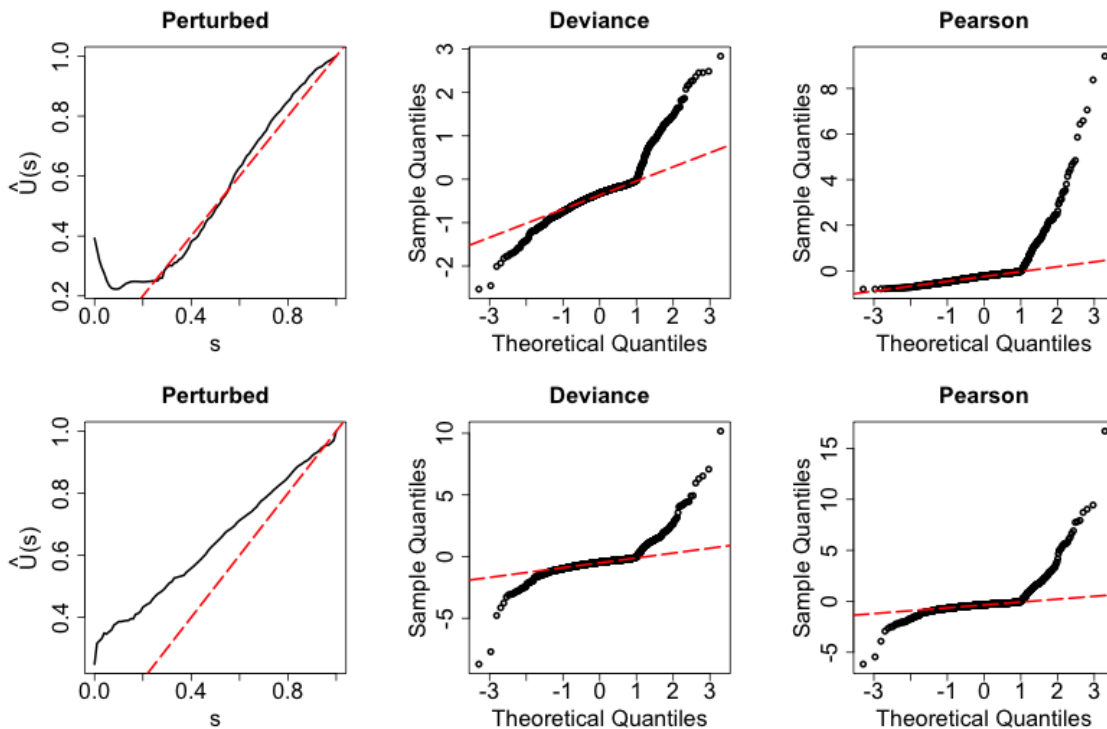
Figure 7.3: Diagnostic plots for negative binomial outcomes under small mean scenario. The two rows correspond to the true model and the misspecified model.

Note that in order to use the proposed method, at least one continuous covariate is necessary. We also address that the tool only provides an informal check on goodness-of-fit. The discrepancy could be due to different causes, for instance wrong choice of link function, wrong choice of covariates, and presence of outliers. If the goodness-of-fit tests suggest

that the model does not fit well overall, the issues can be further checked through specific tests or detecting the influences of individual observations (Cook and Weisberg 1982 and McCullagh and Nelder 1989).

7.2.2 Empirical Study

Recall in Chapter 5, we fit frequencies from building and contents coverage and motor vehicle coverage of the LGPIF with four models: Poisson, negative binomial (NB), zero-inflated Poisson, and zero-one-inflated Poisson models. To choose marginal models, in Chapter 5, we compared empirical and observed counts for each model and used Chi-square goodness-of-fit statistics to compare different models.

Now we do a goodness-of-fit test on each model using the perturbed empirical distribution function as a diagnostic tool. Given the fitted parameters from each model, we compute $\hat{U}(\cdot; \hat{\beta})$ using (2.4). Recall that $\hat{U}(\cdot; \hat{\beta})$ should be close to the identity function under correct model specification.

Figure 7.4 shows the curve of $\hat{U}(\cdot; \hat{\beta})$ for different marginal regression models for BC line. We can see that Poisson and zero-inflated Poisson models fit the data poorly. In contrast, both negative binomial and zero-one-inflated Poisson models give satisfactory results. This is further confirmed by numerical summary in Table 7.1, which display the L_2-norm distances between $\hat{U}(\cdot; \hat{\beta})$ with the 45 degree line

$$d\left(\hat{U}(\cdot; \hat{\beta}), I(\cdot)\right) = \int_s \left(\hat{U}(s; \hat{\beta}) - s\right)^2 \mathrm{d}s,$$

where I is the identity function. We adopt the zero-one-inflated Poisson model here due to its smallest distance, which is consistent with the result depending on goodness-of-fit test statistics in Table 5.5.

For MV line, as in Figure 7.5, the negative binomial model outperforms other models clearly in the sense that the black curve is close to the red dashed line, which is further confirmed numerically in Table 7.1 and consistent with Table 5.7. Here we focus on illustration;

the theoretical properties of the distance will be derived as our future work.
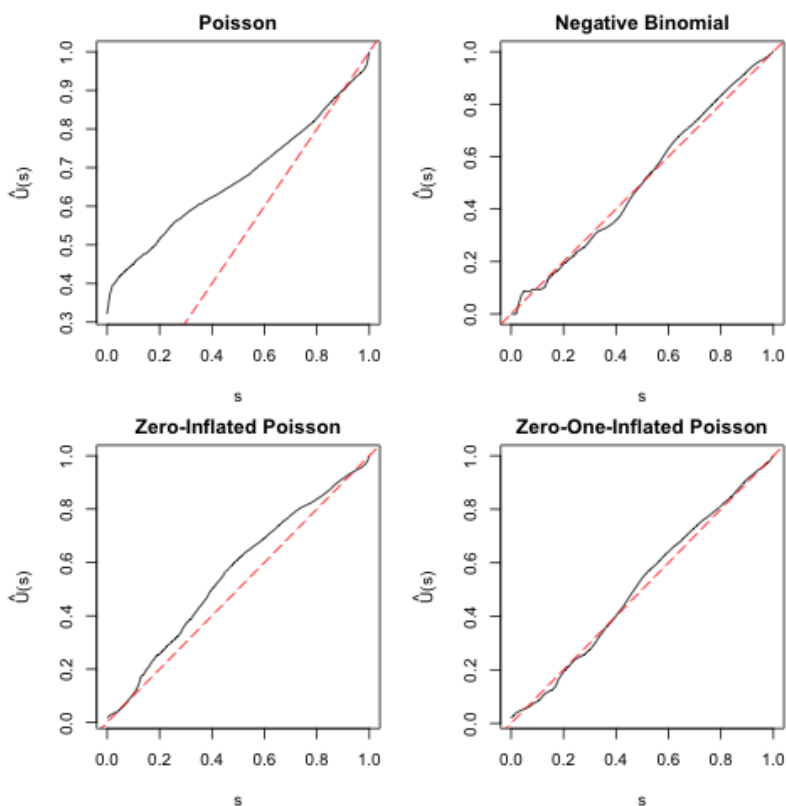
Figure 7.4: Plot of $\hat{U}(\cdot; \hat{\beta})$ (solid line) for marginal models for BC line

Table 7.1: Distance $d\left(\hat{U}(\cdot; \hat{\beta}), I(\cdot)\right)$ from different marginal models (multiplied by 100)

	Poisson	Negative Binomial	Zero-Inflated Poisson	Zero-One-Inflated Poisson
BC	4.799	0.060	0.462	0.057
MV	1.731	0.023	0.685	0.370

7.3 Goodness-of-Fit for Copulas

After choosing and fitting a copula, goodness-of fit tests are necessary for checking how well the copula fits the observations. In this section, we focus on goodness-of-fit tests for copulas on discrete and mixed outcomes. The chapter is structured as follows. In Section
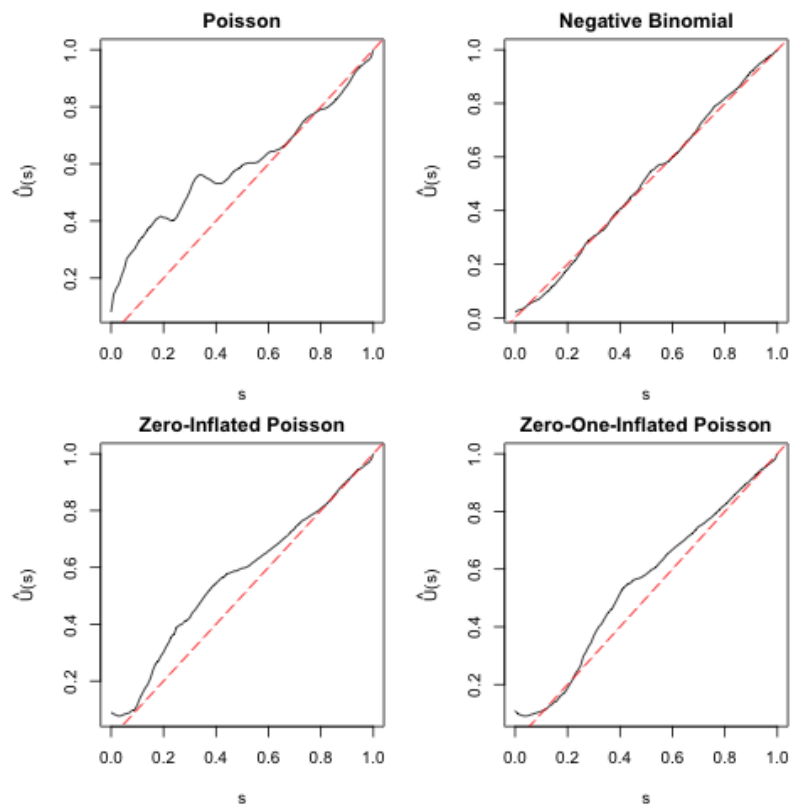
Figure 7.5: Plot of $\hat{U}(\cdot; \hat{\beta})$ (solid line) for marginal models for MV line

7.3.1, we provide a literature review on existing methodologies for continuous and discrete outcomes. In Section 7.3.2 we propose our approach.

7.3.1 Literature

Continuous.

Many goodness-of-fit tests in continuous cases are based on Cox-Snell residuals discussed in Chapter 2. Some formal procedures are summarized in Genest et al. (2009). They mentioned there are other kinds of goodness-of-fit tests for specific copula families but they focused on the "blanket tests" applicable to all copula structures. Using the fact that the empirical distribution function of the Cox-Snell residuals is a consistent estimator of the underlying copula (Deheuvels 1979), hypothesis testing for an assumed copula can be conducted. For example, Cramer-von Mises and Kolmogorov-Smirnov tests can be applied to compare the empirical distribution function of copulas and the fitted parametric copulas.

The Cox-Snell residuals can also be transformed into univariate variables through Kendall's transformation (Genest et al. 2009) which is defined by $\mathbf{U} \to C(U_1, \ldots, U_d)$. With the knowledge of C, the distribution function of the Kendall's transform can be derived. Hence, Cramer-von Mises and Kolmogorov-Smirnov tests can be built to compare the empirical distribution of the Kendall's transforms and their analytical distribution function.

Alternatively, Rosenblatt's transformation through conditioning (Rosenblatt 1952) provides another way of doing model diagnosis. To expand the idea, when multivariate continuous variables $(Y_1, \ldots, Y_d)$ follows a distribution $F(y_1, \ldots, y_d)$, independent uniform random variables can be generated by conditioning

$$
\begin{aligned}
e_1 &= F_1(Y_1) \\
e_2 &= F_{2|1}(Y_2|Y_1) \\
&\ldots \\
e_d &= F_{d|1,\ldots,d-1}(Y_d|Y_1, \ldots, Y_d).
\end{aligned}
\tag{7.1}
$$

Note that the conditional distribution function can be calculated using the marginal models and the copula. The statistic

$$\chi = \sum_{j=1}^{d} \{\Phi^{-1}(e_j)\}^2$$

is supposed to follow a chi-square distribution under the correct model. Hence, Anderson-Darling tests can be applied. Genest et al. (2009) proposed using Cramer–von Mises tests which are based on the empirical distribution function of $(e_1, \ldots, e_d)$ directly.

Overall, their conclusion is that the tests based on Cramer-Von Mises statistics seem better. The Cramer-Von Mises test for empirical distribution and the Cramer-Von Mises test for Rosenblatt's transforms are overall good, although the best depends on the true model and sample size. They also mentioned the order of conditioning in Rosenblatt's transformation requires further research.

Alternatively, Joe (2014) summarized the following methods of goodness-of-fit tests for continuous outcomes:

- Comparison between correlation matrix of normal scores (introduced in Chapter 1) from data and the one from the fitted model. Formula (5.12) in Section 5.8.2 in Joe (2014) gives a goodness-of-fit test statistic.

- Compare empirical versus model-based measures of association such as Kendall's τ and Spearman's ρ.

- Empirical lower and upper semi-correlations of normal scores can be calculated to check the fitting at tail part. An upper tail probability in Section 7.4 in the book can also be used to check goodness-of-fit at upper tail part.

Discrete. For discrete outcomes, goodness-of-fit tests can be done by comparing observed and model-based expected frequencies. Here is the step:

- Collapse the categories of response to avoid zero or small counts

- For each dimension, compare E_j and O_j for univariate goodness-of-fit test, where E_j is the empirical count and O_j is the expected count.

- For pairs, get a two-way table and compare O_{jk} and E_{jk}

The corners of tables from the model-based versus observed bivariate frequencies could be inspected for possible patterns such as over/under-estimation. For example, in Section 7.5 of Joe (2014), in Table 7.12, they find possible lower tail dependence since there is discrepancy among empirical and observed counts at the lower left corner of the table. Nonetheless, the approach of comparing empirical and observed counts is infeasible when there are many large observations, and hard to present when the dimension is greater than 2. In practice, the best model can be selected among candidates by using AIC, BIC, or likelihood. However, these methods are not diagnostic for adequacy of fit and do not suggest improvements. In addition, AIC and BIC do not necessarily indicate goodness of fit at tail.

7.3.2 Proposed Approach

Due to the fact that empirical estimator of Cox-Snell residuals is not a valid estimator of the copula under discreteness, the formal tests summarized in Genest et al. (2009) do not readily apply to discrete outcomes. In this section, as an alternative to the tests based on Rosenblatt's transformation, we propose a way of goodness-of-fit tests using Ruschendorf's transformation (Rüschendorf 2009).

For discrete outcomes, Ruschendorf's transformation uses continuous extensions to obtain uniform variables. That is, to replace the terms in (7.1) by

$$
e_k = \tilde{F}_{k|1,\dots,k-1}(Y_k, V_k | Y_j, j \leq k-1) = F_{k|1,\dots k-1}(Y_k - 1 | Y_j, j \leq k-1)
$$
$$
+ V_k P_{k|1,\dots k-1}(Y_k | Y_j, j \leq k-1),
$$

where $V_1, \dots, V_d$ follow uniform distribution independently, and $P_{k|1,\dots k-1}(\cdot | Y_j, j \leq k-1)$ is the conditional probability. We address that while Ruschendorf's transformation was proposed for discrete outcomes, we can extend it to mixed cases, which could refer the joint distributions of discrete and continuous variables, as well as joint distributions of hybrid variables, i.e., random variables with both discrete and continuous components. Note that

$P_{k|1,\ldots k-1}(\cdot|Y_j, j \le k-1) = 0$ if Y_k is continuous. Hence, we keep the transformation in (7.1) for continuous outcomes. The variables $e_1, \ldots, e_d$ should be uniform if the distribution is specified correctly, which will be checked in Section 7.3.4.

To illustrate, we use the bivariate case as an example. We create two variables,

$$e_1 = F_1(Y_1 - 1) + V_1 P_1(Y_1),$$
$$e_2 = F_{2|1}(Y_2 - 1|Y_1) + V_2 P_{2|1}(Y_2|Y_1). \tag{7.2}$$

The variables e_1 and e_2 should be independently uniformly distributed under correct model specification. If Y_1 is discrete, or for the hybrid outcomes Y takes values from the discrete components,

$$F_{2|1}(y_2|Y_1 = y_1) = \frac{C(F_1(y_1), F_2(y_2)) - C(F_1(y_1 - 1), F_2(y_2))}{P_1(y_1)}.$$

If Y_1 is continuous, or for the hybrid outcomes $Y_1 > 0$, the conditional distribution is

$$F(y_2|Y_1 = y_1) = C_1(F_1(y_1), F_2(y_2)),$$

where C_1 is partial derivative of copula function.

It can be seen that Ruschendorf's transformation is an extension of Rosenblatt's transformation. Therefore, we adapt the corresponding goodness-of-fit tests in Genest et al. (2009) to discrete and mixed cases by applying the fact of independence among $(e_1, \ldots, e_d)$ under correct models.

In bivariate case, let $(e_{i1}, e_{i2}), i = 1, \ldots, n$ be a sample of (e_1, e_2), we apply the probabilistic Kendall's transform

$$H_i = \frac{1}{n} \#\{j : e_{j1} \le e_{i1}, e_{j2} \le e_{i2}\}, \tag{7.3}$$

which should follow distribution function $G(w) = w - w\log(w)$ when e_1 and e_2 are independent uniform variables. Kolmogorov-Smirnov and Cramer-von Mises tests can be used

to test the distribution of H. The empirical quantile of $H_i, i = 1, \ldots, n$ against theoretical quantiles of $G(w)$, named as K-plot in Genest and Boies (2012) can be used to check the model specification. Note that we can also directly apply Kolmogorov-Smirnov and Cramer-von Mises tests on empirical distribution of $\mathbf{e}$ instead of its Kendall's transform. The comparison between these two approaches is part of our future work.

7.3.3 Simulation

Throughout an simulation, we demonstrate the usage of the proposed approach of goodness-of-fit tests on copulas for discrete and mixed outcomes.

We generate the data using a Gaussian copula with parameter 0.5. Marginally, we have the following scenarios

- Discrete (small mean):

  $Y_1 \sim \text{Poisson}(\exp(\beta_{10}+\beta_{11}X_{11}+\beta_{12}X_{12}))$, where $X_{11} \sim N(0,1)$, $X_{12} \sim \text{Bernoulli}(0.4)$, $\boldsymbol{\beta}_1 = (-1,-1,-2)$

  $Y_2 \sim \text{Poisson}(\exp(\beta_{20}+\beta_{21}X_{21}+\beta_{22}X_{22}))$, where $X_{21} \sim N(0,1)$, $X_{22} \sim \text{Bernoulli}(0.7)$, $\boldsymbol{\beta}_2 = (-1,1,-1)$

- Discrete (large mean):

  $Y_1 \sim \text{Poisson}(\exp(\beta_{10}+\beta_{11}X_{11}+\beta_{12}X_{12}))$, where $X_{11} \sim N(0,1)$, $X_{12} \sim \text{Bernoulli}(0.4)$, $\boldsymbol{\beta}_1 = (5,-1,-2)$

  $Y_2 \sim \text{Poisson}(\exp(\beta_{20}+\beta_{21}X_{21}+\beta_{22}X_{22}))$, where $X_{21} \sim N(0,1)$, $X_{22} \sim \text{Bernoulli}(0.7)$, $\boldsymbol{\beta}_2 = (4,1,-1)$

  Figure 7.6 displays the mean levels of discrete outcomes.

- Continuous:

  $Y_1 \sim \text{Gamma}(\exp(\beta_{10} + \beta_{11}X_{11} + \beta_{12}X_{12}))$, where $X_{11} \sim N(0,1)$, $X_{12} \sim \text{Bernoulli}(0.4)$, $\boldsymbol{\beta}_1 = (-1,-1,-2)$, dispersion=3.

$Y_2 \sim \text{Gamma}(\exp(\beta_{20} + \beta_{21}X_{21} + \beta_{22}X_{22}))$, where $X_{21} \sim N(0,1)$, $X_{22} \sim$ Bernoulli(0.7), $\boldsymbol{\beta}_2 = (-1, 1, -1)$, dispersion=2.

- Mixture, which is referred as combination of one continuous and one discrete variable:

  $Y_1 \sim \text{Gamma}(\exp(\beta_{10} + \beta_{11}X_{11} + \beta_{12}X_{12}))$, where $X_{11} \sim N(0,1)$, $X_{12} \sim$ Bernoulli(0.4), $\boldsymbol{\beta}_1 = (-1, -1, -2)$, dispersion=3.

  $Y_2 \sim \text{Poisson}(\exp(\beta_{20}+\beta_{21}X_{21}+\beta_{22}X_{22}))$, where $X_{21} \sim N(0,1)$, $X_{22} \sim$ Bernoulli(0.7), $\boldsymbol{\beta}_2 = (-1, 1, -1)$.

- Tweedie:

  $Y_1 \sim \text{Tweedie}(\exp(\beta_{10} + \beta_{11}X_{11} + \beta_{12}X_{12}))$, where $X_{11} \sim N(0,1)$, $X_{12} \sim$ Bernoulli(0.4), $\boldsymbol{\beta}_1 = (-1, -1, -2)$, dispersion=10, $p = 1.6$.

  $Y_2 \sim \text{Tweedie}(\exp(\beta_{20} + \beta_{21}X_{21} + \beta_{22}X_{22}))$, where $X_{21} \sim N(0,1)$, $X_{22} \sim$ Bernoulli(0.7), $\boldsymbol{\beta}_2 = (-1, 1, -1)$, dispersion=5, $p = 1.4$.

We do the simulation with 500 replications and 1000 as sample size. For each of the replications, we fit different copulas: Gaussian, Frank, and Clayton copulas. Then we compute (e_{i1}, e_{i2}) using (7.2) and obtain the Kendall's transforms using (7.3). We then conduct the Kolmogorov-Smirnov test for the distribution of the Kendall's transforms. For comparison, we also include the case in which the copula parameter is under/overestimated as 0.2 and 0.8. Table 7.2 shows average p-values of the 500 replications. For each copula, the first column shows p-value for Y_1 and $Y_2|Y_1$, and the second column shows the one for Y_2 and $Y_1|Y_2$.

The conclusion is that first, we can make correct conclusion for continuous outcomes to reject wrong models, except that it is hard to exclude the Frank copula due to its similarity to the Gaussian copula. Second, for discrete outcomes, especially when the mean is small and the data are distributed on a few points, it is hard to distinguish among different copulas as all the models are not rejected. When the mean gets larger, we are able to reject copulas when the tail dependence structure are different, for example, Clayton copulas have
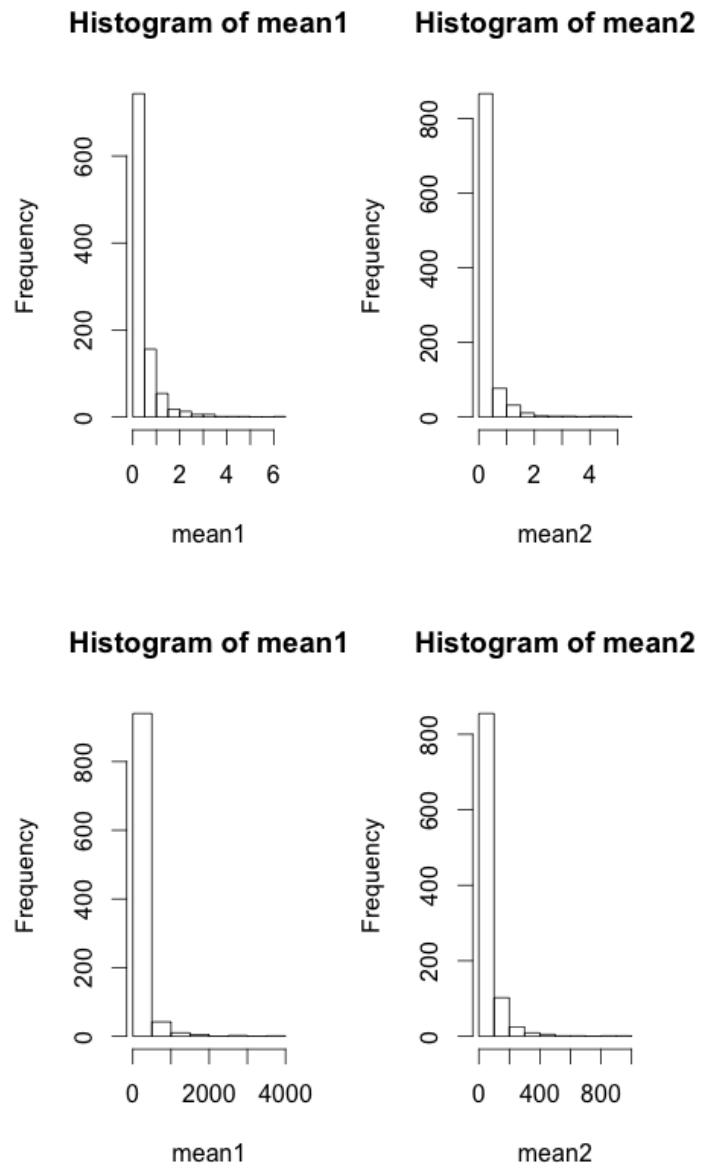
Figure 7.6: Mean of one replication for discrete outcomes. Upper: small mean. Lower: large mean.

Table 7.2: Average p-values of 500 replications

	Gaussian		Gaussian(0.8)		Gaussian(0.2)		Frank		Clayton	
	p	p	p	p	p	p	p	p	p	p
Poisson (small mean)	0.637	0.610	0.230	0.236	0.281	0.329	0.627	0.612	0.587	0.594
	(0.251)	(0.256)	(0.224)	(0.228)	(0.226)	(0.248)	(0.247)	(0.268)	(0.259)	(0.271)
Poisson (large mean)	0.661	0.657	0.00000	0.00000	0.00000	0.00000	0.375	0.364	0.058	0.059
	(0.243)	(0.235)	(0.00000)	(0.00000)	(0.00001)	(0.00002)	(0.230)	(0.235)	(0.075)	(0.077)
Gamma	0.557	0.507	0.00001	0.00001	0.00000	0.00000	0.307	0.286	0.050	0.053
	(0.281)	(0.280)	(0.0001)	(0.00005)	(0.00002)	(0.00002)	(0.225)	(0.225)	(0.071)	(0.085)
Mixture	0.640	0.612	0.042	0.027	0.052	0.044	0.631	0.603	0.048	0.488
	(0.262)	(0.262)	(0.063)	(0.056)	(0.075)	(0.073)	(0.260)	(0.257)	(0.060)	(0.253)
Tweedie	0.632	0.616	0.257	0.249	0.436	0.437	0.634	0.618	0.635	0.618
	(0.253)	(0.261)	(0.234)	(0.234)	(0.244)	(0.244)	(0.258)	(0.264)	(0.258)	(0.264)

lower tail dependence while Gaussian copulas do not have tail dependence. Even for large mean case, it is hard to distinguish between Gaussian and Frank copulas, which happens for all the scenarios. This is consistent with our conclusions in Section 4.3. Joe also made comments that "it is typical for multivariate discrete data that copula models with similar dependence structure provide approximately the same level of fit because the effects of the tail behavior of the copulas is attenuated" (Page 327 in Joe 2014). Third, we cannot make conclusion for Tweedie outcomes in our simulations. This is due to the influence of large proportion of zero. In our setting, we have 80% zeros on average. We expect it becomes easier to pick up the correct copula when we have less zeros. Finally, it is noticeable that the order of conditioning plays an important role for the mixture case.

7.3.4 Check of Independence

Below, we check the independence of the Rosenblatt's and Ruschendorf's transforms.

Continuous.

$$
\begin{aligned}
P(e_1 \leq u_1, e_2 \leq u_2) &= \int_{-\infty}^{\infty} P(F_{2|1}(Y_2|Y_1 = x) \leq u_2, F_1(Y_1) \leq u_1|Y_1 = x) f_1(x) \mathrm{d}x \\
&= \int_{-\infty}^{F_1^{-1}(u_1)} P(F_{2|1}(Y_2|Y_1 = x) \leq u_2) f_1(x) \mathrm{d}x \\
&= \int_0^{u_1} u_2 \mathrm{d}x = u_1 u_2
\end{aligned}
$$

where the third equation is due to and $F_{2|1}(Y_2|Y_1 = x)$ follows a uniform distribution. Hence, e_1 and e_2 independently follow uniform distribution.

Discrete.

$$
P(e_1 \leq u_1, e_2 \leq u_2) = \int_0^1 \int_{-\infty}^{\infty} P(\tilde{F}_{2|1}(Y_2, V_2|Y_1 = x) \leq u_2, \tilde{F}_1(Y_1) \leq u_1|Y_1 = x, V_1 = v_1) f_1(x) \mathrm{d}x \mathrm{d}v_1
$$

$$(7.4)$$

Since $\tilde{F}_1(x|V_1 = v_1) = F_1(x - 1) + v_1 P_1(x)$ is a monotone increasing function in x. We can

take inverse with respect to x conditioning on v_1, denote as $\tilde{F}_1^{-1}(u_1|v_1)$. Therefore,

$$P(e_1 \leq u_1, e_2 \leq u_2) = \int_0^1 \int_{-\infty}^{\tilde{F}_1^{-1}(u_1|v_1)} P(\tilde{F}_{2|1}(Y_2, V_2|Y_1 = x, V_1 = v_1) \leq u_2|Y_1 = x, V_1 = v_1) f_1(x) \mathrm{d}x \mathrm{d}v_1$$

$$= u_2 \int_0^1 \int_{-\infty}^{\tilde{F}_1^{-1}(u_1|v_1)} f_1(x) \mathrm{d}x \mathrm{d}v_1 = u_1 u_2$$

Mixture. Assume Y_1 is discrete while Y_2 is continuous.

$$P(e_1 \leq u_1, e_2 \leq u_2) = \int_0^1 \int_{-\infty}^{\tilde{F}_1^{-1}(u_1|v_1)} P(F_{2|1}(Y_2|Y_1 = x, V_1 = v_1) \leq u_2|Y_1 = x, V_1 = v_1) f_1(x) \mathrm{d}x \mathrm{d}v_1$$

$$= u_2 \int_0^1 \int_{-\infty}^{\tilde{F}_1^{-1}(u_1|v_1)} f_1(x) \mathrm{d}x \mathrm{d}v_1 = u_1 u_2$$

Bibliography

Aas, Kjersti, Claudia Czado, Arnoldo Frigessi, and Henrik Bakken (2009) "Pair-copula constructions of multiple dependence," *Insurance: Mathematics and Economics*, Vol. 44, pp. 182–198.

Acar, Elif F, Radu V Craiu, and Fang Yao (2011) "Dependence calibration in conditional copulas: A nonparametric approach," *Biometrics*, Vol. 67, pp. 445–453.

Agresti, Alan and Maria Kateri (2011) *Categorical Data Analysis*: Springer.

Aitchison, John and CH Ho (1989) "The multivariate Poisson-log normal distribution," *Biometrika*, Vol. 76, pp. 643–653.

Bedford, Tim and Roger M Cooke (2002) "Vines–a new graphical model for dependent random variables," *The Annals of Statistics*, Vol. 30, pp. 1031–1068.

Berkhout, Peter and Erik Plug (2004) "A bivariate Poisson count data model using conditional probabilities," *Statistica Neerlandica*, Vol. 58, pp. 349–364.

Bermúdez, Lluís and Dimitris Karlis (2011) "Bayesian multivariate Poisson models for insurance ratemaking," *Insurance: Mathematics and Economics*, Vol. 48, pp. 226–236.

Brown, Charles E (1998) "Multivariate probit analysis," in *Applied Multivariate Statistics in Geohydrology and Related Sciences*: Springer, pp. 167–169.

Chen, Song Xi and Tzee-Ming Huang (2007) "Nonparametric estimation of copula functions for dependence modelling," *Canadian Journal of Statistics*, Vol. 35, pp. 265–282.

Chib, Siddhartha and Rainer Winkelmann (2012) "Markov chain Monte Carlo analysis of correlated count data," *Journal of Business & Economic Statistics*.

Chiu, Shean-Tsong (1991) "Bandwidth selection for kernel density estimation," *The Annals of Statistics*, pp. 1883–1905.

Cook, R Dennis and Sanford Weisberg (1982) *Residuals and influence in regression*: New York: Chapman and Hall.

Cox, David R and E Joyce Snell (1968) "A general definition of residuals," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 248–275.

Czado, Claudia, Ulf Schepsmeier, and Aleksey Min (2012) "Maximum likelihood estimation of mixed C-vines with application to exchange rates," *Statistical Modelling*, Vol. 12, pp. 229–255.

Deheuvels, Paul (1979) "La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance," *Acad. Roy. Belg. Bull. Cl. Sci.(5)*, Vol. 65, pp. 274–292.

Denuit, Michel and Philippe Lambert (2005) "Constraints on concordance measures in bivariate discrete data," *Journal of Multivariate Analysis*, Vol. 93, pp. 40–57.

Fan, Jianqing and Irène Gijbels (1992) "Variable bandwidth and local linear regression smoothers," *The Annals of Statistics*, pp. 2008–2036.

——— (1995) "Adaptive order polynomial fitting: bandwidth robustification and bias reduction," *Journal of Computational and Graphical Statistics*, Vol. 4, pp. 213–227.

Frees, Edward W (1995a) "Assessing cross-sectional correlation in panel data," *Journal of Econometrics*, Vol. 69, pp. 393–414.

——— (1995b) "Semiparametric estimation of warranty costs," *Journal of Nonparametric Statistics*, Vol. 5, pp. 103–122.

Frees, Edward W, Richard A Derrig, and Glenn Meyers (2014) *Predictive modeling applications in actuarial science*, Vol. 1: Cambridge University Press.

Frees, Edward W, Xiaoli Jin, and Xiao Lin (2013) "Actuarial applications of multivariate two-part regression models," *Annals of Actuarial Science*, Vol. 7, pp. 258–287.

Frees, Edward W, Gee Lee, and Lu Yang (2016) "Multivariate frequency-severity regression models in insurance," *Risks*, Vol. 4, p. 4.

Frees, Edward W and Emiliano A Valdez (1998) "Understanding relationships using copulas," *North American Actuarial Journal*, Vol. 2, pp. 1–25.

Genest, Christian and Jean-Claude Boies (2012) "Detecting dependence with Kendall plots," *The American Statistician*.

Genest, Christian and Johanna Nešlehová (2007) "A primer on copulas for count data," *Astin Bulletin*, Vol. 37, pp. 475–515.

Genest, Christian, Aristidis K Nikoloulopoulos, Louis-Paul Rivest, and Mathieu Fortin (2013) "Predicting dependent binary outcomes through logistic regressions and meta-elliptical copulas," *Brazilian Journal of Probability and Statistics*, Vol. 27, pp. 265–284.

Genest, Christian, Bruno Rémillard, and David Beaudoin (2009) "Goodness-of-fit tests for copulas: A review and a power study," *Insurance: Mathematics and economics*, Vol. 44, pp. 199–213.

Haff, Ingrid Hobæk, Kjersti Aas, and Arnoldo Frigessi (2010) "On the simplified pair-copula construction–Simply useful or too simplistic?" *Journal of Multivariate Analysis*, Vol. 101, pp. 1296–1310.

Hansen, Bruce E (2009) "Lecture notes on nonparametrics."

Joe, Harry (1993) "Parametric families of multivariate distributions with given margins," *Journal of Multivariate Analysis*, Vol. 46, pp. 262–282.

———— (1997) *Multivariate Models and Multivariate Dependence Concepts*: CRC Press.

———— (2014) *Dependence Modeling with Copulas*: CRC Press.

Johnson, Norman Lloyd, Samuel Kotz, and Narayanaswamy Balakrishnan (1997) *Discrete Multivariate Distributions*, Vol. 165: Wiley New York.

Karlis, Dimitris and Loukia Meligkotsidou (2005) "Multivariate Poisson regression with covariance structure," *Statistics and Computing*, Vol. 15, pp. 255–265.

Kolev, Nikolai and Delhi Paiva (2009) "Copula-based regression models: A survey," *Journal of Statistical Planning and Inference*, Vol. 139, pp. 3847–3856.

Krämer, Nicole, Eike C Brechmann, Daniel Silvestrini, and Claudia Czado (2013) "Total loss estimation using copula-based regression models," *Insurance: Mathematics and Economics*, Vol. 53, pp. 829–839.

Li, Bo and Marc G Genton (2013) "Nonparametric identification of copula structures," *Journal of the American Statistical Association*, Vol. 108, pp. 666–675.

Li, David X (1999) "On default correlation: A copula function approach," *Available at SSRN 187289*.

Li, Yan, Yang Li, Yichen Qin, and Jun Yan (2016) "Copula modeling for data with ties," *arXiv preprint arXiv:1612.06968*.

McCullagh, Peter and John A Nelder (1989) *Generalized Linear Models*, Vol. 37: CRC press.

McCulloch, Charles E and John M Neuhaus (2001) *Generalized Linear Mixed Models*: Wiley Online Library.

Nelsen, Roger B (2003) "Properties and applications of copulas: A brief survey," in *Proceedings of the First Brazilian Conference on Statistical Modeling in Insurance and Fi-*

nance,(Dhaene, J., Kolev, N., Morettin, PA (Eds.)), *University Press USP: Sao Paulo*, pp. 10–28, Citeseer.

——— (2007) *An Introduction to Copulas*: Springer Science & Business Media.

Nikoloulopoulos, Aristidis K (2013) "Copula-based models for multivariate discrete response data," in *Copulae in Mathematical and Quantitative Finance*: Springer, pp. 231–249.

Nikoloulopoulos, Aristidis K and Dimitris Karlis (2008) "Multivariate logit copula model with an application to dental data," *Statistics in Medicine*, Vol. 27, pp. 6393–6406.

——— (2009) "Modeling multivariate count data using copulas," *Communications in Statistics-Simulation and Computation*, Vol. 39, pp. 172–187.

——— (2010) "Regression in a copula model for bivariate count data," *Journal of Applied Statistics*, Vol. 37, pp. 1555–1568.

Ohlsson, Esbjörn and Björn Johansson (2006) "Exact credibility and Tweedie models," *Astin Bulletin*, Vol. 36, pp. 121–133.

Omelka, Marek, Irene Gijbels, Noël Veraverbeke et al. (2009) "Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing," *The Annals of Statistics*, Vol. 37, pp. 3023–3058.

Panagiotelis, Anastasios, Claudia Czado, and Harry Joe (2012) "Pair copula constructions for multivariate discrete data," *Journal of the American Statistical Association*, Vol. 107, pp. 1063–1072.

Patton, Andrew J (2006) "Modelling asymmetric exchange rate dependence," *International Economic Review*, Vol. 47, pp. 527–556.

Pierce, Donald A and Daniel W Schafer (1986) "Residuals in generalized linear models," *Journal of the American Statistical Association*, Vol. 81, pp. 977–986.

Randles, Ronald H (1984) "On tests applied to residuals," *Journal of the American Statistical Association*, Vol. 79, pp. 349–354.

Rosenblatt, Murray (1952) "Remarks on a multivariate transformation," *The Annals of Mathematical Statistics*, Vol. 23, pp. 470–472.

Ruppert, David, Simon J Sheather, and Matthew P Wand (1995) "An effective bandwidth selector for local least squares regression," *Journal of the American Statistical Association*, Vol. 90, pp. 1257–1270.

Rüschendorf, Ludger (2009) "On the distributional transform, Sklar's theorem, and the empirical copula process," *Journal of Statistical Planning and Inference*, Vol. 139, pp. 3921–3927.

Scaillet, Olivier and Jean-David Fermanian (2002) "Nonparametric estimation of copulas for time series," *FAME Research paper*.

Shi, Peng (2016) "Insurance ratemaking using a copula-based multivariate Tweedie model," *Scandinavian Actuarial Journal*, Vol. 2016, pp. 198–215.

Shi, Peng and Edward W Frees (2011) "Dependent loss reserving using copulas," *Astin Bulletin*, Vol. 41, pp. 449–486.

Shi, Peng and Emiliano A Valdez (2014) "Multivariate negative binomial models for insurance claim counts," *Insurance: Mathematics and Economics*, Vol. 55, pp. 18–29.

Shi, Peng and Lu Yang (2017) "Pair copula constructions for insurance experience rating," *Journal of the American Statistical Association*.

Shih, Joanna H and Thomas A Louis (1995) "Inferences on the association parameter in copula models for bivariate survival data," *Biometrics*, pp. 1384–1399.

Sklar, M (1959) *Fonctions de Répartition À N Dimensions et Leurs Marges*: Université Paris 8.

Smyth, Gordon K (1989) "Generalized linear models with varying dispersion," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 47–60.

Song, Peter X-K, Mingyao Li, and Ying Yuan (2009) "Joint regression analysis of correlated data using Gaussian copulas," *Biometrics*, Vol. 65, pp. 60–68.

Song, Peter Xue-Kun (2007) *Correlated Data Analysis: Modeling, Analytics, and Applications*: Springer Science & Business Media.

Sukhatme, Balkrishna V (1958) "Testing the hypothesis that two populations differ only in location," *The Annals of Mathematical Statistics*, pp. 60–78.

Tsionas, Efthymios G (2001) "Bayesian multivariate Poisson regression," *Communications in Statistics-Theory and Methods*, Vol. 30, pp. 243–255.

Vuong, Quang H (1989) "Likelihood ratio tests for model selection and non-nested hypotheses," *Econometrica: Journal of the Econometric Society*, pp. 307–333.

Winkelmann, Rainer (2000) "Seemingly unrelated negative binomial regression," *Oxford Bulletin of Economics and Statistics*, Vol. 62, pp. 553–560.

Yang, Xipei, Edward W Frees, and Zhengjun Zhang (2011) "A generalized beta copula with applications in modeling multivariate long-tailed data," *Insurance: Mathematics and Economics*, Vol. 49, pp. 265–284.

Young, Gary, Emiliano A Valdez, and Robert Kohn (2009) "Multivariate probit models for conditional claim-types," *Insurance: Mathematics and Economics*, Vol. 44, pp. 214–228.

Zeger, Scott L and Kung-Yee Liang (1986) "Longitudinal data analysis for discrete and continuous outcomes," *Biometrics*, pp. 121–130.

Zhang, Yanwei (2013) "Likelihood-based and bayesian methods for tweedie compound poisson linear mixed models," *Statistics and Computing*, Vol. 23, pp. 743–757.